National Defence / Défense nationale

MODULATION RECOGNITION: AN OVERVIEW

by

René Lamontagne

DEFENCE RESEARCH ESTABLISHMENT OTTAWA

TECHNICAL NOTE 91-3

Canada

March 1991
Ottawa

National Defence / Défense nationale

MODULATION RECOGNITION: AN OVERVIEW

by

René Lamontagne
Communications Electronic Warfare Section
Electronic Warfare Division

DEFENCE RESEARCH ESTABLISHMENT OTTAWA

TECHNICAL NOTE 91-3

PCN
041LK11

March 1991
Ottawa

ABSTRACT

The interest for a system able to automatically identify the modulation type of an intercepted radio signal is increasingly evident for military and civilian purposes. Although in the past some authors looked at the problem, nobody found "the solution" and the problem remains. Therefore, an overview of the proposed techniques is useful in order to assess the actual situation.

This document presents classification techniques and features (parameters characterizing the modulation types) used for modulation recognition.

RÉSUMÉ

Que ce soit pour des fins militaires ou civiles, l'intérêt d'une machine capable d'identifier automatiquement le type de modulation d'un signal inconnu est évident. Bien que par le passé certains auteurs se soient penchés sur le problème, personne n'a obtenue d'éclatant résultats indiquant la route à suivre. Par conséquent il n'y a pas lieu de concentrer ses efforts sur un seul auteur mais sur l'ensemble des techniques proposées.

Ce document couvre donc le sujet d'une façon globale, présentant les techniques pour classification ainsi que les paramètres utilisés pour caractériser et distinguer les divers types de modulation, tel que proposés par les auteurs.

EXECUTIVE SUMMARY

The motivation of this document is the interest of military and civilian organisations in monitoring the electromagnetic signal activity in the RF spectrum. Since the number of competent trained human operators is constantly decreasing and the radio activity in the HF and VHF bands increasing, the interest of a machine capable of automatically identifying the modulation type of an unknown intercepted signal is quite obvious. Integrating this device into an ESM system including energy detection (spectral analysis) Direction Finding and Data Fusion and Correlation, would allow an operator to drastically improve his efficiency and his ability to monitor the activity in the RF spectrum.

Although in the past some authors looked at the problem and proposed algorithms permitting to achieve proper performance for high SNR signals, for an ESM perspective low SNR signals are more likely to be intercepted. Therefore, the problem of finding good parameters able to discriminate among the modulation types of interest when the noise is important, still remains and is very realistic. More and more publications appear in the literature presenting new ideas and better performance. Also, with the appearance of better hardware processors, new possibilities are now available and it is believed that soon a modulation recognition device will be able to classify very noisy signals in a short period of time with a high accuracy. Although such a device is not yet available, the work done until now is certainly worthy and merits consideration.

This document introduces and presents techniques proposed in the open literature for automatic modulation type identification of an intercepted radio signal. Most of these techniques are based on the same classic pattern recognition theory, which is presented in a separate section, since it is a prerequisite to understand the following sections. The two main classification techniques used for modulation recognition, the linear classifier and the decision tree, are more specifically discussed.

Modulation recognition is concerned with analog as well as digital modulation types. Although there is a modern tendency to replace analog modulation types by digital ones, analog modulations (i.e., SSB, FM, AM, etc.) are still in use in many countries. The motivation to intercept these signals is reinforced by the fact that these signals, once identified, can also be demodulated to extract the message. It might not be the case with digital modulation types, since usually coding is involved: error correcting code, encryption, vocoder, etc.

Since the main differences between the approaches proposed by the referred authors are the features they used, the main part of the document consists in presenting these features and the corresponding results. Also, more particular points are considered such as preprocessing to remove gaps in analog amplitude modulated signals.

The purpose of this technical note is to summarize in a very comprehensive way all the publications on the topic of modulation recognition.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

1.0 INTRODUCTION

From the earliest days of radio communication the need to monitor the electromagnetic signal activity in the RF spectrum has existed. Civilian authorities may wish to monitor the transmissions over their territory in order to maintain a control over this activity. Military organizations may wish to monitor the radio activities of other powers for reasons, among others, of national security.

Typically, the technique employed by monitoring stations throughout the world is based upon a well–tried method which has been in use for many years. This is a one man–one receiver situation in which the operator spends his time searching the RF spectrum with a continuously tunable general purpose receiver, hoping to make an interesting interception. There are variations to that scenario. For example, it is possible and common practice for an operator to have two receivers. One is used for searching, while the second remains tuned to a known interesting frequency. A further variation is called the master–slave technique. In this method a group of operators is involved. The master operates a fast tuning–sweep receiver with an associated panoramic display unit. When a signal of interest is located, this intercepted signal is transferred to one of the slaves, who tunes to the appropriate frequency and does the monitoring with a conventional, general purpose, continuously tunable receiver.

Neither of these techniques overcomes the fundamental problem of serious overcrowding in the radio spectrum. Moreover, the existing pool of highly skilled operators has begun to dry up and it is proving difficult to find replacements [1]. The classical method of monitoring is after all, a very boring occupation. Therefore Electronic Support Measure (ESM) techniques become an important alternative.

In advanced ESM systems, the operator is helped or replaced by sophisticated electronic machines. These machines are concerned with exploiting enemy electromagnetic emissions for the purpose of gathering intelligence information as automatically as possible. This information is provided by analysis of the attributes of an intercepted signal. Thus, Modulation Recognition (MR) is an ESM technique: given an intercepted signal, it aims to identify the modulation type among a number of known possible modulation types. The terms *modulation classification*, *recognition* or *identification* are currently used to describe this process.

Prior to modulation classification, the radio signal must be intercepted, which means that somewhere before the modulation classification system, there is an energy detection system looking for electromagnetic emissions in the bandwidth of interest. Once a signal has been detected, the logical following step is to try to identify this signal. A step farther is the demodulation of that signal and then decryption to finally obtain the signal itself.

Thus classification is neither energy detection nor normal signal demodulation with message extraction; it is something in between (see Figure 1).

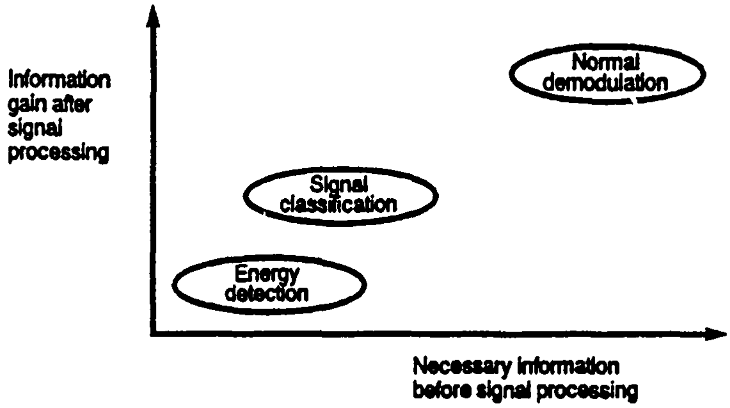

Figure 1: Informational Relationships [2, p.312]

For energy detection, only the bandwidth of interest for the ESM system is known. On the other hand, for demodulation with message extraction, knowledge of the center frequency, bandwidth, type of modulation, data rate...parameters is required. The signal classifier should need only the information given by the energy detection system, i.e. where the signal is: center frequency and bandwidth.

An example of an ESM system is illustrated in Figure 2. The intercepted signal is submitted to energy detection algorithms, then down–converted and bandpass filtered, prior to modulation recognition. The information obtained from the latter and from the energy detector is gathered by the system controller, which will assign the signal to a proper demodulator, as shown in Figure 2.

The particular problems related to modulation classification are caused by the radio channel and the ESM system components. The potential problems could be summarized by:

Effect	Source
–Multipath fading	Radio Channel
–Poor SNRs	Radio Channel
–Amplitude distortion	Receiver amplifier
–More than one signal	Energy Detector
–Incomplete signal	Energy Detector
–Frequency instability	Down Convertor

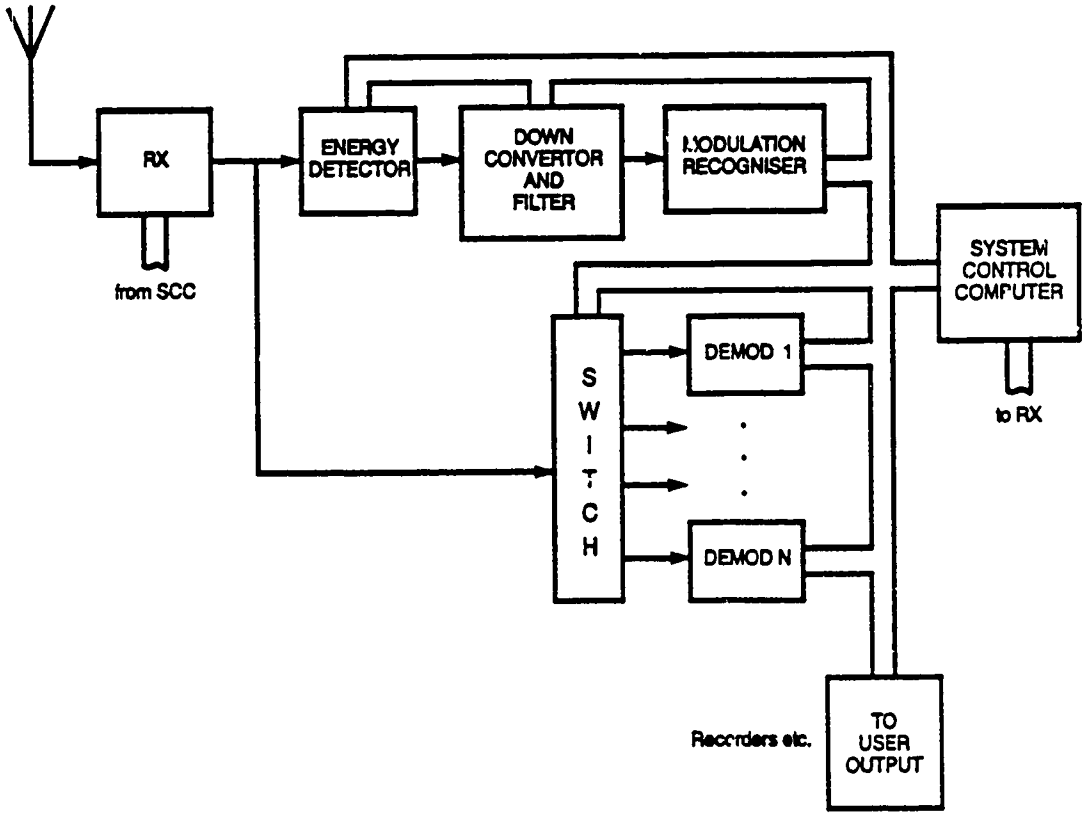

Figure 2: Example of a simple ESM system.

The multipath fading is not discussed in the literature, but the problem of poor SNR is. Although classifiers for high SNRs exist, those for low SNRs are rare. According to some simulations [3][4][5][6], under ideal circumstances, it would be possible to classify a signal when the SNR is above 8 dB (results for a real system and real signals could be different). All the classifiers presented in the literature assume only one signal in the passband of the narrowband filter. Thus it is assumed that the energy detection algorithm is able to separate every signal perfectly. This is not trivial. The frequency stability is also a problem raised in the literature [7]. The performance of the classifier should not be too affected by mistuning the down–converter. Hipp [8] got acceptable performance with carriers within 10% of the typical value.

Furthermore, modulation recognition presents problems which are due to the nature of the signal itself. This can be summarized as follows:

Problem	Source
–insufficient signal	short acquisition time
–unmodulated segments	gaps in the voice (analog modulations)
–long computing time	too many features

The performance of a classifier is closely related to the quantity of information available. Thus it is possible to improve the performance of a classifier with longer acquisition times (more sample points) and/or by computing more features from the signal. In both cases the classification process time is increased. The acquisition time is especially long for analog modulations. AM, SSB, DSB and FM are afflicted with "unmodulated segments" caused by gaps in the voice. These gaps result in CW segments for AM and FM, and in noise segments for SSB and DSB. Gallant [3] presented a technique to remove these unmodulated segments at the price of an acquisition time of at least 1.5 sec. For digital modulation types, the problem is to get enough symbol transitions, otherwise CW will be detected. The problem is especially obvious at low data rates. For example, the data rate of OOK could be as low as 10 Hz, then five seconds are required to get only 50 transitions. The effect of the acquisition time on the system performance can be perceived in [6], [9] and [10].

The computing time depends mainly upon the number of sample points, the number of extracted features, the number of modulation types, and the complexity of the classification algorithm. These features permit the user to discriminate among the modulations. More features could provide more discriminating facilities. Also, features are usually time invariant, which means they are computed over all the sample points. For these reasons, the features should be easy to compute, so that a reasonable overall computing time can be obtained. Also, although some highly sophisticated classification algorithms exist, the ones used in pattern recognition are simple. The overall computing time presented in the literature is about 2 seconds.

In the following sections the process of modulation recognition itself will be treated. Firstly, the concept of *classification* or *recognition* will be presented. Although they describe the same problem, the term *classification* and *recognition* do not represent exactly the same reality. *Recognition* (from pattern recognition) includes an additional step before classification: feature extraction and/or selection. This is preparing the data for classification. Two kinds of classifiers are presented and used for modulation recognition: the linear classifier and the decision tree. They are widely used in real systems because they are simple and fast enough for real–time applications. Finally, the researchers in the area of MR will be summarized in a matrix, in preparation for the next sections which present features used by them in their Modulation Recognition (MR) systems.

Section 3.0 presents MR systems capable of recognizing analog modulation types. The features useful for that purpose are highly sensitive to noise for an obvious reason: analog modulation schemes consist mainly of amplitude modulation (AM, SSB, DSB). However, a few authors presented "less sensitive" features. Another topic introduced in this section is the problem of gaps in the voice. We will see in detail the interesting solution proposed by Gallant.

The features for digital modulation schemes (Section 4.0) are quite different. Although the separation between Sections 3.0 and 4.0 is very arbitrary (most authors do a few of both), different features are required to be able to get information on the phase of phase modulated signals (BPSK, QPSK, FSK). These features are given in Section 4.0.

The two preceding sections present features used by pattern recognition algorithms for the purpose of modulation recognition. A few authors proposed an alternative way of doing MR, by using energy detection algorithms. This perspective is presented in section 5.0. Unfortunately, the performance expected from these techniques is not discussed by the authors.

Finally, Section 6.0 concludes the document with a short summary and comments.

2.0 CLASSIFICATION TECHNIQUES

2.1 GENERAL

The term *classification* is the action of associating individuals into one of two or more alternative *classes* (groups) on the basis of a set of inputs called *features* (variables). The populations are known to be distinct according to the features. As an example, consider an archeologist who wishes to determine which of two possible tribes created a particular statue found in a dig. The archeologist takes measurements for several characteristics of the statue and decides which tribe these measurements are most likely to have come from. The measurements of the statue may consist of a single observation such as its height, however, we would then expect a low degree of accuracy. If on the other hand the classification is based on several characteristics, we would have more confidence in the prediction.

Classification algorithms could be used in almost any area of knowledge, it is the heart of any decision process. Tou [11] divided problems where classification algorithms are applied into two major categories:

1. The study of human beings and other living organisms,

2. The development of theory and techniques for the design of devices capable of performing a given recognition task for a specific application.

The first subject area is concerned with such disciplines as sociology, psychology, physiology and biomedical sciences. The second area is concerned with computer, and engineering aspects of the design of automatic *pattern recognition* systems. Pattern recognition can be defined as the categorization of input data into identifiable classes via the extraction of significant features followed by a classification process. Thus in pattern recognition we are talking of a two step process: *feature extraction* and *classification*. Contrary to the preceding examples, in pattern recognition there are no direct features available. Although data is available under a digital form (signal processing ADC), more processing is required to give some meaning to this data. During the feature extraction process, the large quantity of data is translated into a few significant and discriminant features used by the classifier.

Pattern recognition spans a number of disciplines and problems, as shown in the following list:

–speech recognition	words identification
–speaker recognition	speaker identification
–speaker verification	speaker identification (knowing the words used)
–character recognition	character identification
–visual inspection	object anomaly identification
–ship recognition	kind of ship identification
–biomedical analysis	medical diagnoses
–weather prediction	weather forecast
–stock market prediction	predicted market ups and downs

Early pattern recognition research performed in the '60s and '70s focused on the asymptotic (infinite training data) properties of classifiers. Many researchers studied *parametric* Bayesian classifiers, where the form of input distributions is assumed to be known, and parameters of distributions are estimated using techniques that require simultaneous access to all training data. These classifiers, especially those that assume Gaussian distributions, are still the most widely used because they are simple and described in a number of textbooks.

The thrust of recent research has changed, much of it motivated by the desire to understand and build parallel neural net classifiers inspired by biological neural networks. This has led to an emphasis on robust, adaptive, *non-parametric* classifiers that can be implemented on parallel hardware. It is very likely that future modulation recognition systems will use this new technology.

In this chapter, some classical classification techniques will be presented. These techniques are used in pattern recognition as well as in human sciences. The concept of feature extraction and selection will also be introduced. Then the particular problem of MR itself will be presented.

2.2 PATTERN RECOGNITION TECHNIQUES

The goal of pattern recognition is to assign input patterns to one of k classes. The input patterns consist of static input vectors x containing n elements (continuous or discrete values) denoted x_1, x_2, x_3,..., x_n. These elements represent measurements of features selected to be useful for distinguishing between classes and insensitive to irrelevant variability in the input. A good classification performance requires the selection of effective features as well as the selection of a classifier that can make good use of those features with limited training data, memory, and computing power. During the *training phase*, a limited amount of training data and *a priori* knowledge concerning the expected output is used to adjust parameters and/or learn the structure of the classifier. Once the training is accomplished, the classifier is ready for the *test phase*, during which a new set of inputs is presented to the classifier without *a priori* knowledge (see Figure 3). The performance is then computed and presented, usually in a confusion table (percentage of good classification for each class).

The subject of feature selection and extraction is concerned with reducing the dimensionality of pattern representation. Since the complexity of a classifier grows rapidly with the number of dimensions of the pattern space, it is important to base decisions only on the most essential, so-called discriminatory information. Dimensionality reduction is also recommended from a classification performance point of view. Initially performance improves as new features are added, but at some point, inclusion of further features will result in performance degradation.

Dimensionality reduction can be achieved in two different ways. One approach is to identify measurements which do not contribute significantly to class separability. this is feature selection. The other approach, called feature extraction, consists of mapping the useful information in a lower-dimension feature space (see Figure 4).

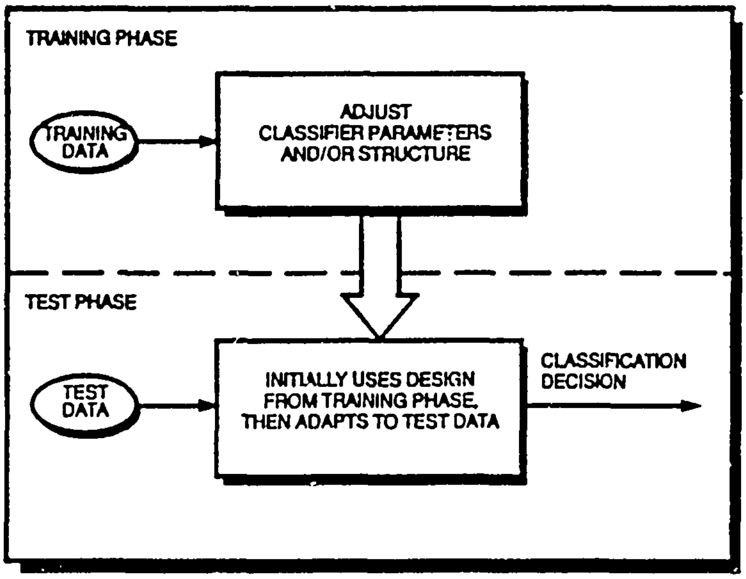

Figure 3: The two major phases for pattern classifier development[12, p.48]

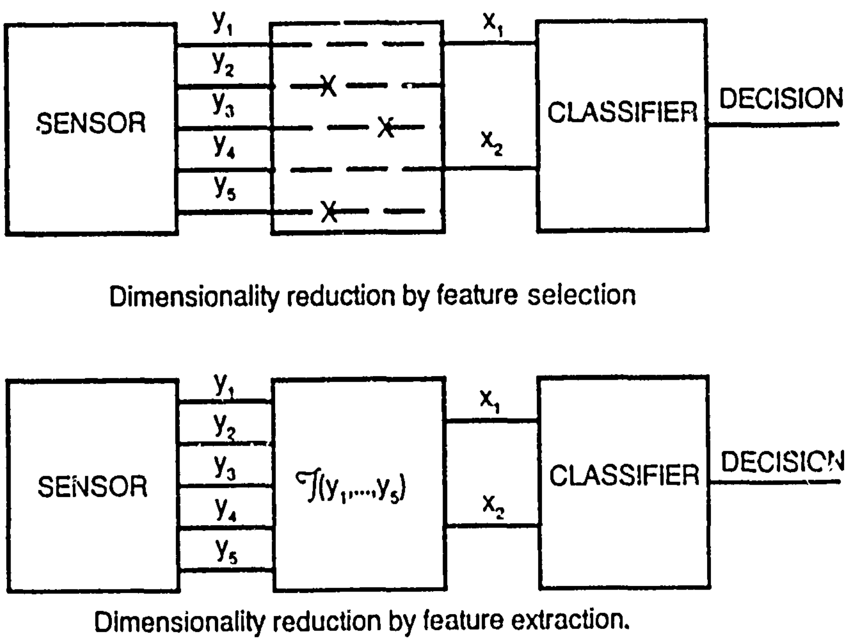

Figure 4. Dimensionality reduction by feature extraction/selection.

To solve a feature selection and/or feature extraction problem, we need some sort of evaluation criterion. Unfortunately this is not a trivial problem. The quality of a set of features is very closely related to the classifier used. Therefore the optimal procedure is to try all the possible sets of features with the classifier and to retain the "best" set. In general this process requires too much computation and therefore a number of alternative feature evaluation criteria exist.

It is not the purpose of this document to explain the details of feature evaluation criteria, however the basic idea will be given in Sections 2.2.1 and 2.2.2.

2.2.1 Feature Selection

We will now briefly discuss how to evaluate a feature by using the following example. Suppose that an individual may belong to one of two populations. We begin by considering how an individual can be classified into one of these populations on the basis of a measurement of one characteristic, say X. We have a representative sample of this measure from each population. The distribution is represented in Figure 5.

From the figure above, it is obvious that a good feature must have a small variance among the samples and a mean highly discriminative among classes. Ideally the distribution should not overlap so that there is no misclassification. If the two distributions have the same variance and same prior probabilities, the decision rule is quite trivial, and the threshold is at the intersection of the distribution.

Combining more variables (or features) may provide better classification accuracy. Consider two variables, X_1 and X_2, with distributions similar to X. By combining the two variables according to a linear function ($Z = a_1X_1 + a_2X_2$), we get a two dimensional region (see Figure 6). Usually Z is called a canonical function.

The example presented in Figure 6 shows that by linearly combining some features, a decision region is created in which there are clusters. The samples are gathered in clusters according to similar patterns. In the classical approach discussed here, there is one cluster per class. As shown in Figure 6, by adding a feature, the clusters should be more distinct, with less overlapping (in the dimension N). Thus the feature selection criterion should evaluate the distance between clusters and verify the amount of overlapping. There are numerous ways to evaluate the distance, the two most common being probably the Mahalanobis and Euclidean methods. For a complete list of distance measurement criteria, see [13].

Unfortunately, even with distinct and non–overlapping clusters, the classifier may not provide the performance expected according to a feature evaluation criterion. Depending upon the shape of the clusters, some classifiers may not be able to adequately divide the decision region (this will be explained in Section 2.2.3). Therefore the real evaluation criterion is done by a stepwise analysis, trying the desired classifier with the features, adding and removing them, and comparing the results.

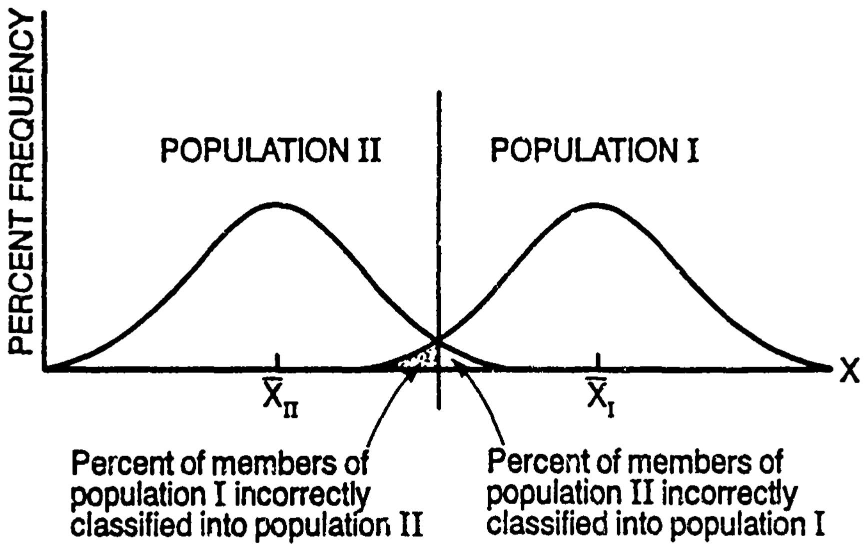

Figure 5: Distribution of the feature for the two classes.

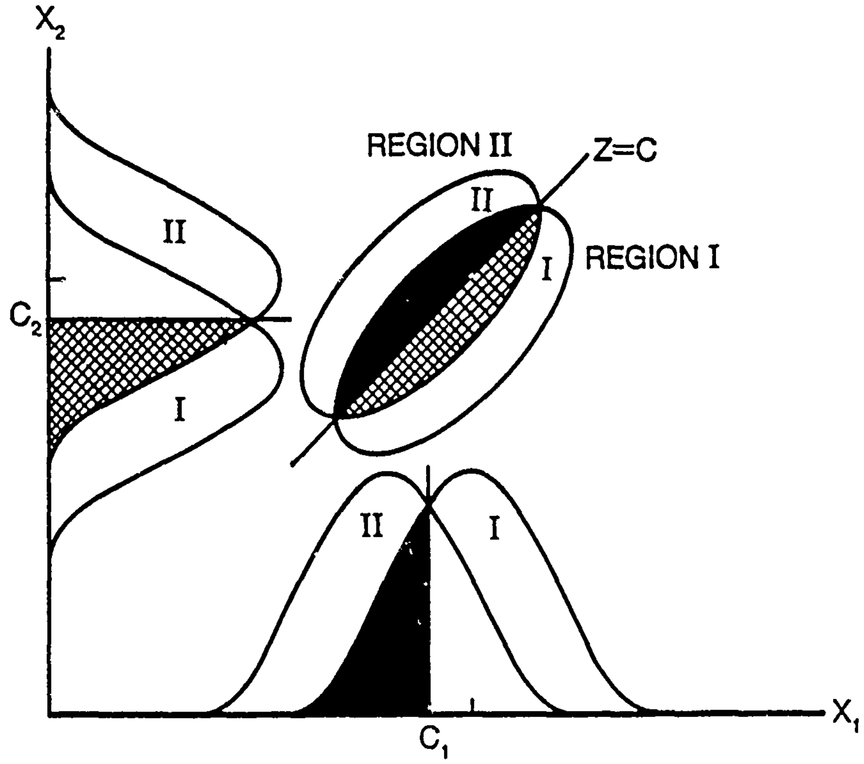

Figure 6: Decision Region: 2 features and 2 classes.

2.2.2 Feature Extraction

As pointed out in Section 2.2, in feature extraction all the features are used in order to create a lower–dimensional space, thus reducing the classifier complexity. Information compression is achieved by a mapping process in which all the useful information contained in the original observation vector y is converted onto a few composite features of vector x, while ignoring redundant and irrelevant information.

$$x = \mathcal{F}(y)$$

$$y = [y_1, y_2, \ldots, y_n]^T$$

$$x = [x_1, x_2, \ldots, x_m]^T$$

$$n > m$$

Although non–linear mapping is possible, linear mapping is much more usual and has the advantage of being computationally feasible. The mapping function becomes a matrix multiplication.

$$y = Tx$$

$$T = [t_1, t_2, \ldots, t_n]$$

t_i are column vectors

$T^T T = I$: T is orthonormal

x is obtained using

$x_i = t_i^T y$, for i=1, 2,..., m

This kind of compression technique is called Principal Component Analysis (PCA) in the statistical literature [14, pp.309–330], and the Karhunen–Loève Expansion in pattern recognition literature [15, pp.226–250], [16].

PCA can be summarized as a method of transforming the original variables into new uncorrelated variables to avoid redundancy. The new variables are called the *principal components.* Each principal component is a linear combination of the original variables. The principal components are chosen to keep the mean–square error between y and yy minimal, where yy, the estimation for y, is defined by:

$$yy = \sum_{1}^{m} x_i t_i + \sum_{m+1}^{n} b_i t_i$$

where b_i are preselected constant.

The matrix T is computed according to the eigenvectors of the covariance (or autocorrelation) matrix of the distributions of the y_i [15, p.236].

2.2.3 Classification Techniques

The concern in this section involves the determination of optimum decision procedures, which are needed in the identification process. After the observed data from patterns to be recognized have been expressed, a machine is required to decide to which class ω_i these data belong.

By and large the only generally valid statistical decision theory is based upon the average cost or loss in misclassification, formulated in terms of the Bayes expressions for conditional probabilities (briefly called "Bayes classifier"). In standard pattern recognition theory it is reasonably accurate to assume that the unit misclassification cost is the same for all classes. Assuming that x is the vector of input observations (pattern elements, sets of attributes ...) and $\{\omega_i,\ i = 1, 2,..., k\}$ is the set of classes to which x may belong, let $p(x|\omega_i)$ be the probability density function of x in class ω_i, and $P(\omega_i)$ be the *a priori* probability of occurrence of samples from class ω_i; in other words, $d_i(x) = p(x|\omega_i)P(\omega_i)$ corresponds to the class distribution of those samples of x which belong to class ω_i. Here the $d_i(x)$ are called the discriminant functions. The average rate of misclassification is minimized if x is conclusively classified according to the following rule:

$$x \text{ is assigned to } \omega_i \text{ iff } d_i(x) > d_j(x),\ \forall\, j \neq i$$

The main problem, of course, is to obtain analytic expressions for the $d_i(x)$. Notice that even a large number of samples of x, as such, does not define any analytical probability density function. One has to use either *parametric* or *nonparametric* methods.

Parametric (also called probabilistic) methods assume *a priori* probability distributions (such as Gaussian) for input features. Parameters of distributions (means, variances, covariances,...) are estimated using supervised training where all data is assumed to be available simultaneously. These classifiers provide optimal performance when the underlying distributions are accurate models of the test data and sufficient training data is available to estimate the distribution parameters accurately. Although these two conditions are not necessarily satisfied with real–world applications, these classifiers are popular since they are simple and sufficiently efficient in many cases. In the literature on MR, the most common choice is Fisher's [17] linear classifier presented in Section 2.2.3.1.

Although nonparametric techniques exist such as the *k–nearest neighbor classifier*, they are not popular classifiers for MR, being too complex and time consuming. They also require huge amounts of training data. One exeption is the *binary tree classifier* (also called decision tree, classification tree,...). Since it is used for MR, it will be presented in Section 2.2.3.2.

2.2.3.1 Linear Classifiers

It is assumed that a pattern vector $x = [x_1, x_2,..., x_m]^T \in \omega_i$, $i \leq k$ (w_i are the possible classes) is presented to a classifier. As shown in Figure 6, it is possible to draw a straight line between them and call this line the decision boundary, threshold, or discriminant function $d(x)$.

$$d(x) = W_0 + W_1x_1 + W_2x_2$$

$$d(x) > 0 \rightarrow x \in \omega_1$$
$$d(x) < 0 \rightarrow x \in \omega_2$$

For the general case of multiclasses, there are as many discriminant functions as classes. Then the decision is taken according to the rule

$$x \text{ is assigned to } \omega_i \text{ iff } d_i(x) > d_j(x), \forall j \neq i$$

Therefore we can write

$$d_1(x) = W_0 + W_1x_1 + W_2x_2$$

$$d_2(x) = -W_0 - W_1x_1 - W_2x_2$$

Usually, in pattern recognition, the matricial notation is preferred. A new vector, z, is introduced

$$z = [1, x_1, x_2, \ldots, x_m]^T$$

$$W_1 = [W_0, W_1, \ldots, W_m]^T$$
$$W_2 = [-W_0, -W_1, \ldots, -W_m]^T$$

$$d_1 = W_1^Tz$$
$$d_2 = W_2^Tz$$

$$d = W^Tz$$

$$d = [d_1, d_2, \ldots, d_k]^T$$

$$W = [W_1, W_2, \ldots, W_k]$$

Figure 7 shows a few possible cluster shapes. It is clear in the figure that a linear classifier cannot efficiently discriminate some clusters. It is uncertain that a linear classifier will be able to determine proper linear functions when there are several clusters. However other kinds of classifiers are available, such as the *quadratic classifier.*

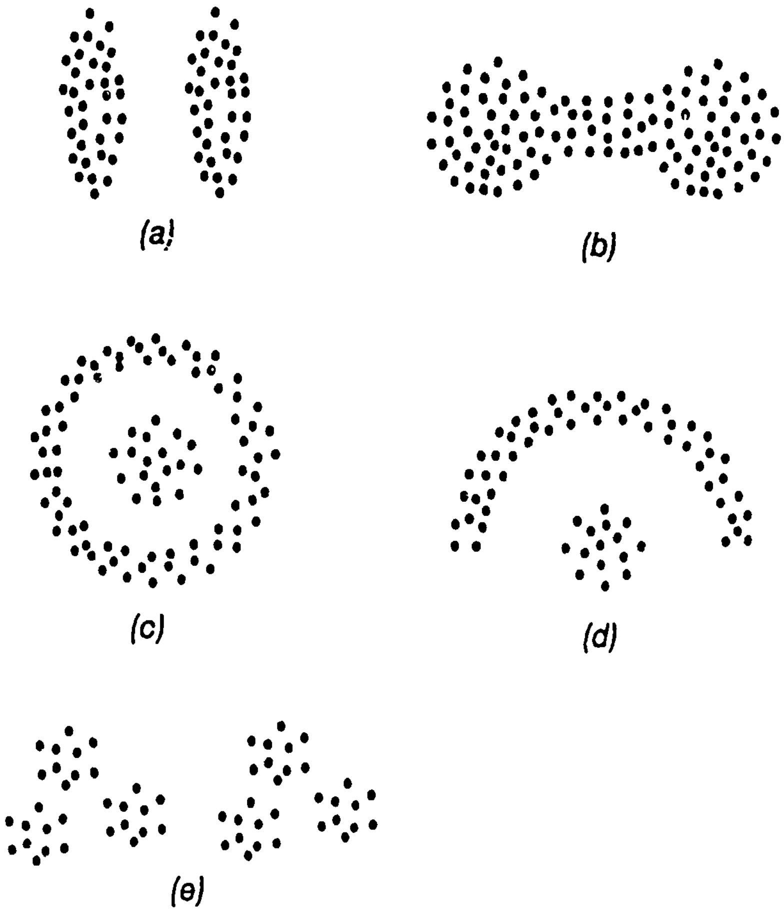

Figure 7: Cluster shapes possible (two classes) [18, p.35].

a) Compact and well–separated clusters.
b) Touching clusters.
c) Concentric clusters.
d) Linearly nonseparable clusters.
e) Multi–modal clusters.

The linear classifier was expressed as

$$\mathrm{d} = W^T z$$

$$\mathrm{d} = W^T x + c,$$

c being the thresholds, i.e. W_0.

The quadratic classifier is expressed as

$$\mathrm{d} = y^T V x + W^T x + c$$

The boundaries are curves instead of straight lines. Then clusters as in Figure 7–d can be properly discriminated, however, the computing and memory requirements are significantly higher. In some situations, the performance improvement obtained with the quadratic classifier is so small that it does not justify its complexity.

2.2.3.2 Decision Tree Classifiers

Decision tree classifiers are hyperplane classifiers which have been developed extensively over the past 10 years. It is a rather different method of discriminant analysis which portrays the problem in terms of a binary tree. The tree provides a hierarchical–type representation of the data space that can be used for classification by tracing up the tree.

The line of development started in 1963 and has attracted growing interest in the last 10 years, developing a large number of algorithms able to create binary trees and multiple binary trees. For the latter see [19]. A classical technique, called CART, will be presented here.

In its simplest form, the CART [20] method produces a tree based upon individual variables. For example, the split at the bottom of the tree might be determined by the question, "Is $x_5 < 6.2$?". This will determine a left and right branch. The left branch corresponding to $x_5 < 6.2$ might then be divided according to the question, "Is $x_3 > 1.4$?" and the right branch for which $x_5 > 6.2$, might be split according to the question, "Is $x_1 > 0$?". The methodology has three components: the set of questions, the rules for selecting the best splits and the criterion for choosing the extent of the tree. With the tree trained, each terminal node of the tree is associated with one of the class ω_i.

More sophisticated questions can also be handled, such as, "Is $\Sigma W_i x_i <$ Threshold?". Numerous questions are possible and can be mixed in the same tree. Although the concept is very simple, the implementation of efficient algorithms able to optimize the tree is not: at each node the algorithm must be able to select the best question and the best feature, and must know when to terminate the tree. Note also that it is a nonparametric procedure requiring complex and efficient training algorithms, as well as a considerable amount of training data. This alternative is especially interesting when a lot of classes are involved and a lot of features are available.

2.3 MODULATION RECOGNITION

The preceding techniques will now be applied to solve the problem of Modulation Recognition (MR). As introduced in the first chapter, the signal presented to the MR system is intercepted by a receiver, down–converted and bandpass filtered (see Figure 8).

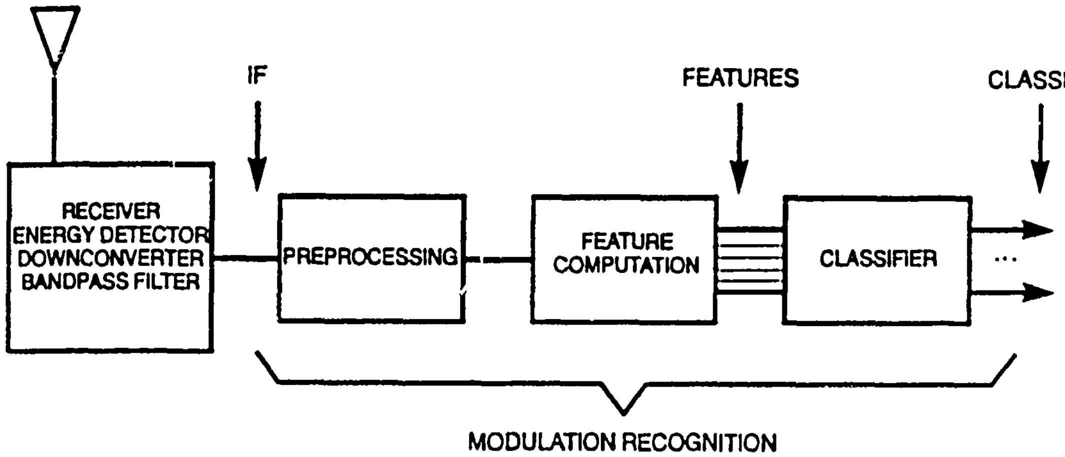

Figure 8: General Architecture of a MR system

The intercepted signal is usually derived from a computer–controlled intercept receiver. For example, Gallant [3] cites the Watkins–Johnson WJ–8607 receiver [21]. The RF spectrum is scanned to find a signal. When a signal is detected we have a rough idea of its frequency. Thus the signal can be down–converted to an intermediate frequency which is usually fixed. Thus, the output of the receiver is an IF signal ready for the classifier

The preprocessing stage varies a lot from author to author, and can even be nonexistant. An example of preprocessing is shown in Gallant [3]. He uses a quite complex three–step preprocessing scheme to remove the "unmodulated" part of the AM, FM, DSB and SSB signals. These "unmodulated" parts are caused by gaps in the voice, mainly between words.

The preprocessing stage can be either analog or digital. Although most of the authors used DSP boards, some preferred analog preprocessing. For example, Winkler [22] sent the IF signal to a bank of parallel analog demodulators, one for each modulation type. Calian (previously Miller) [23]–[27] used three analog PLL demodulators, one each for AM, FM and DSB.

To allow some versatility and flexibility, feature computation is accomplished digitally. Given an intercepted signal, a feature is a signal characteristic useful to discriminate among the possible modulation types. These features should be as discriminative as possible, even when the noise level is significant. Most of the classifiers do not perform well at low SNRs.

The decision procedures considered by the authors are represented by the two techniques explained in Section 2.2: the linear and the decision tree classifiers. The decision tree technique considered here consists of the simplest form: questions of the type "Is $x_4 > 3.2$?", where 3.2 would have been obtained during the training phase. Moreover the tree is not optimized: instead of using a sophisticated algorithm like CART to establish the optimal splits and thresholds, the sets of rules are defined empirically by looking at the training data. Note finally that the tree is sometimes presented in alternative ways: boolean equations or logic table (see Figure 9).

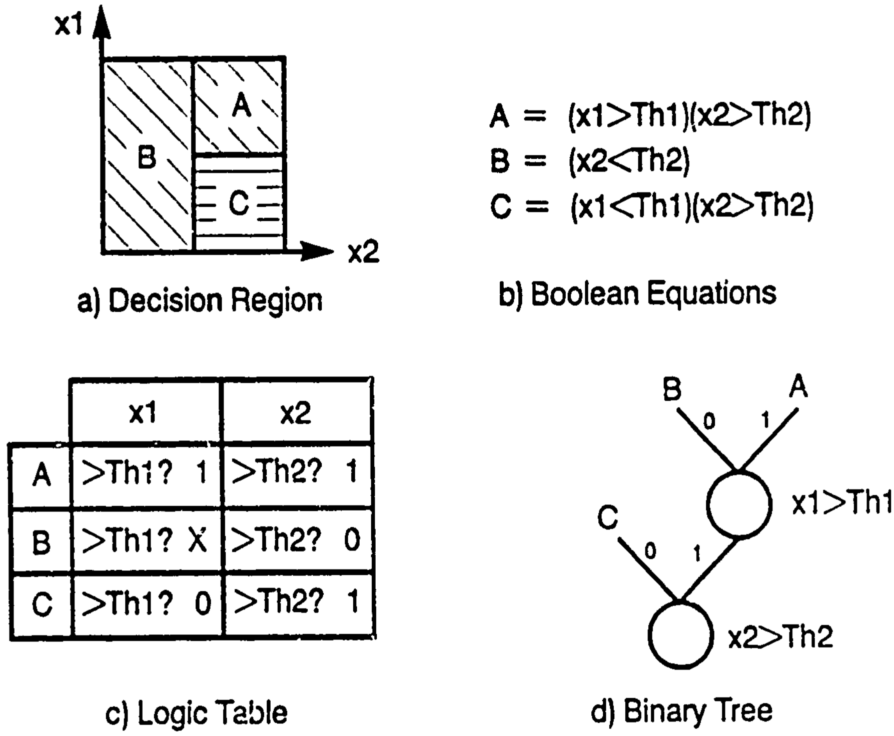

Figure 9: Illustration of the logic tree decision process.

Usually the authors, especially the latest ones, will use a linear classifier. However one author, Jondral [28][29], got slightly better results with a quadratic classifier.

To collect the information on MR, a literature search has been conducted. All the papers reported here are unclassified. They report an interest in recognizing both analog and digital modulations: AM, SSB, DSB, FM, ASK,OOK, BPSK, QPSK and FSK. From a military perspective, Torrieri [30] in his book says that the importance of analog communication systems is declining, probably due to the proliferation of digital computers and the security provided by cryptographic digital communication. Thus in the last years more papers have been published on digital modulation classifiers. Nevertheless, analog communications are still in use and the problem of analog modulation classification still retains some interest.

The techniques for modulation classification described in the following pages have been grouped depending on whether the author put more emphasis on analog or digital modulation schemes. Within each group, the authors have been gathered according to some similarities, as shown in Table 1. In the table, "?" refers to a modulation type for which it was not clear whether or not the MR algorithm was able to recognize.

GROUP	AUTHOR	SECTION	MODULATION TYPE								
			ANALOG				DIGITAL				
			AM	SSB	DSB	FM	ASK	FSK	BPSK	QPSK	OTHER
ANALOG	Miller	3.1	✓	✓	✓	✓			✓		
	Luiz		✓	✓	✓	✓			✓		
	Wakeman		✓	✓	✓	✓			✓		
	Fry		✓	✓	✓	✓					
	Gadbois	3.2	✓	✓	✓	✓		✓	✓		
	Ribble		✓	✓	✓	✓	✓	✓	✓		
	UTL		✓	✓	✓	✓		✓	✓		
	Gallant		✓	✓	✓	✓					
	Weaver	3.3	✓	✓				✓			
	Winkler		✓	✓	✓	✓					
	Callaghan		✓	✓		✓	?	?			
	Fabrizi		✓	✓		✓					
	Petrovic		✓	✓		✓	✓	✓			
	Aisbett	3.4	✓		✓	✓					
	Einicke		✓	✓	?	✓					
	Hipp	3.5	✓	✓	✓	✓	✓	✓	✓		
DIGITAL	Liedtke	4.1					✓	✓	✓	✓	4-FSK
											8-PSK
	Jondral		✓	✓			✓	✓	✓		4-FSK
	Dominguez		✓	✓	✓	✓	✓	✓	✓	✓	4-ASK
											4-FSK
	Adams										
	Mammone	4.2							✓	✓	
	DeSimio						✓	✓	✓	✓	
ENERGY DETECTOR	Ready	5.1						?	?	?	
	Kim	5.2							✓	✓	
	Gardner	5.3									

Table 1: Groups of authors.

3.0 MR FOR ANALOG MODULATIONS

This group is bigger for a historical reason: the first paper was written in 1969. The interest in digital modulation is more recent. The particularity of this group is the features used. Because analog modulations are dominated by amplitude modulation (AM, DSB, SSB), the signal envelope is a characteristic somewhat exploited by several authors, starting with Gadbois in 1985. Before that, authors were using a hardware approach as we will see with Miller.

3.1 MILLER APPROACH

3.1.1 Miller

In 1978, Miller Communications Systems Limited (now Calian Communications Systems Limited) presented a report [23] to DREO on the feasibility of a HF/VHF spectrum surveillance receiver, including automatic modulation type identification. DREO liked the idea and prepared a contract for a prototype ESM receiver. The approach developed by Miller for modulation classification prevailed for years and led to a number of publications [23]–[27][31][32][7].

In their implementation, the receiver produces a 455kHz IF signal (with a passband filter of bandwidth either 6 or 24 kHz) which is fed to an envelope sampler APD (Amplitude Probability Distribution) and three parallel phase–lock–loop (PLL) demodulator circuits (see Figure 10).

The method used to characterize the signal amplitude envelope is called APD (Amplitude Probability Distribution). As shown in Figure 10, the output of the envelope detector is sampled by an 8 bit ADC (with a sampling rate of 1.7kHz). N points were used to compute the following statistics:

$$\text{Mean: } \mu = \frac{1}{N} \Sigma (x)$$

$$\text{Variance: } s^2 = \frac{1}{N} [\Sigma (x^2) - \frac{1}{N} (\Sigma x)^2]$$

P (x): Probability that the ADC output is equal to x, corresponding to the APD (see Figure 11).

After an inspection of 2^{13} samples, Miller found that good statistical significance was obtained and could be useful for modulation classification. However, to keep the number of features low, only two APD parameters have been retained: the variance s^2 and the probability that the output of the ADC equals 0, P (0).

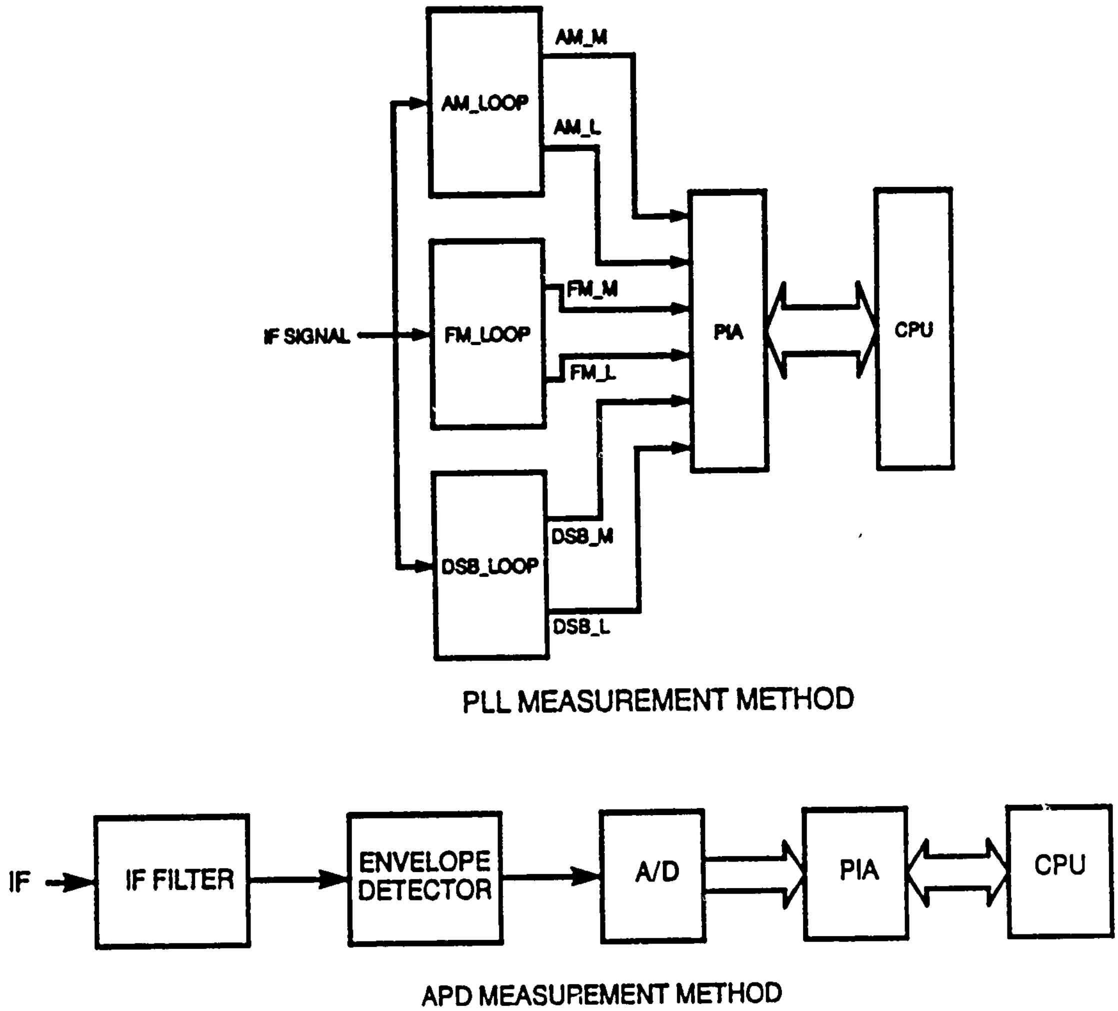

Figure 10: Miller's classifier [25]

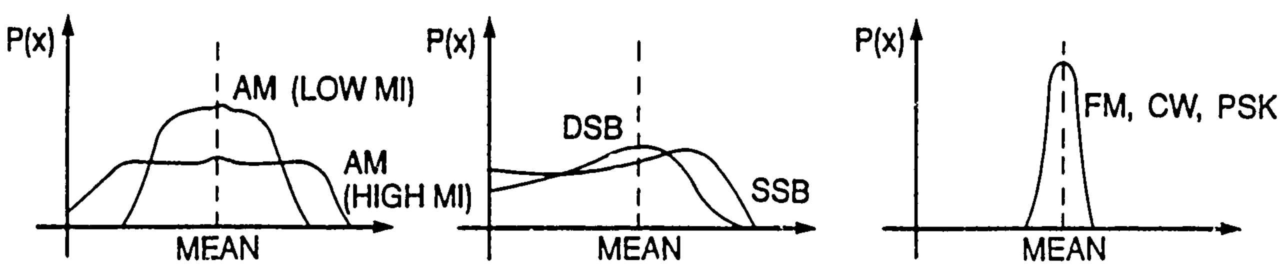

Figure 11: Theoretical APDs [25].

In a similar way, Miller studied the output of the PLL circuits in order to use the property that different types of PLL circuits have a tendency to lock onto different signals. Each of the three loops (AM, FM and DSB) has 2 outputs used for the classification: a Lock indicator and a Modulation indicator. Considering 4096 samples, Miller created a table which is reproduced in Table 2. This table shows which PLL is locked to which modulation type.

Combining both APD and PLL results, Miller created a *Reference Logic Table* (see Table 2). The table contains eight binary features: two from the APD and six from the PLL. The binary values are the results of hard decisions on the outputs of the ADCs with reference thresholds.

SIGNAL	AM-M	AM-L	FM-M	FM-L	DSB-M	DSB-L	S^2	P(0)
AM	1	1	0	1	1	1	1	0
FM	X	0	1	1	X	0	0	0
PSK	1	0	1	1	1	1	0	0
CW	0	1	0	1	0	1	0	0
DSB	X	0	1	0	1	1	1	1
SSB	X	0	X	1	X	0	1	1
Threshold	.312	.812	.625	.500	.250	.812	256	.004

Table 2: Miller's Reference Logic Table [25].

The classification is accomplished by comparing the unknown with the reference table for a perfect fit. To use the terminology introduced in Section 2, it is a very simple *binary tree classifier*. The bandwidth of the passband filter is first set at 24 kHz. If there is no match for the classification, another attempt is made with the 6 kHz filter.

The results obtained by Miller are quite impressive. They claim a percentage of correct identification higher than 90% for all modulation types. However, the SNR is not specified. To get these results, the classifier used 256 samples per try and a minimum of four tries per classification.

As stated previously, the Miller prototype ESM receiver was built under a DND contract. An evaluation of Miller's work is reported by Luiz in [31].

3.1.2 Wakeman

In 1985, a U.S. Patent was registered [32] for a "Spectrum Surveillance Receiver System". It is Miller's receiver, no upgrade to the classification procedure has been reported.

3.1.3 Fry

Fry [7] also looked at Miller's receiver. Some problems had been reported[1], and Fry was assigned by DREO to determine their origins. After some tests, he found that due to the nature of the PLL design used, the accuracy of the identification process deteriorated rapidly if the receiver was not tuned to the center of the signal. He also noted that identification failed with real voice signals. Tests with the modified prototype (modification both in the hardware and decision tree) revealed 95% correct identification with a 10dB SNR, when a sine wave was used as modulating signal. Unfortunately, the accuracy fell drastically for real voice signal. With an infinite SNR (no noise) and a real voice signal (normal gaps associated with continuous speech), he got an average performance around 60%.

3.2 GADBOIS APPROACH

3.2.1 Gadbois

In 1985, Gadbois [33]–[35] introduced a novel alternative to modulation identification, based on the envelope characteristics of the received signal. The feature selected is the ratio R of the variance to the square of the mean of the squared envelope. Intuitively, since FM has constant envelope while AM does not, the former's R is zero and the latter's is close to unity. After some studies Gadbois found that "...SSB, DSB, AM and FM have very distinctive R's...". He developed mathematical relations relating R to the SNR, as presented in Figure 12. These expressions are based on the assumption of having infinite record lengths. In the case of samples, the estimation of R, noted $\hat{R}$, is computed as followed.

$$\sigma^2 \approx \hat{\sigma}^2 = \frac{1}{N}\,\Sigma\,(x^2) - \frac{1}{N^2}\,(\Sigma x)^2$$

$$\mu \approx \hat{\mu} = \frac{1}{N}\,\Sigma x$$

$$R \approx \hat{R} = \frac{\hat{\sigma}^2}{\hat{\mu}^2} = N\,\frac{\Sigma\,(x^2)}{(\Sigma x)^2} - 1$$

---

1. "Laboratory use of the receiver revealed three major shortcomings: speed, accuracy and friendliness.", and talking about modulation classification, "For some reason, the unit did not work well and its performance fluctuated from day to day.".

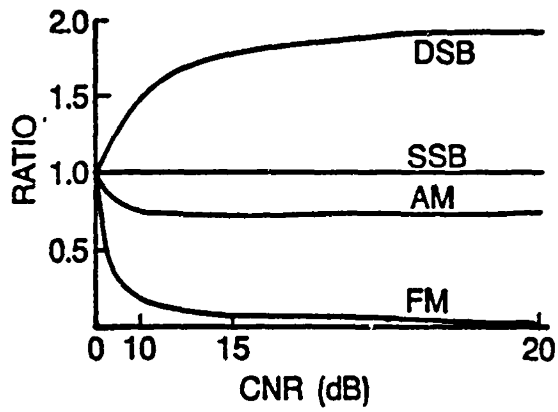

Figure 12: Mathematical relations relating R to the SNR [35, p.152].

Computing $\hat{R}$ for N=2048 points per sample for a large number of samples, Gadbois determined threshold values within a small decision tree for classification of AM, SSB, DSB and FM. In his experiments, the voice was simulated by Gaussian white noise low–pass filtered. The sampling rate for the 12–bit ADC was 160kHz. The IF frequency was 40kHz. The results of that very simple classifier were computed from 200 samples per modulation type and are presented here in Table 3.

ACTUAL	CLASSIFIED AS			
	FM	AM	SSB	DSB
FM	200	0	0	0
AM	0	181	19	0
SSB	0	15	160	25
DSB	0	0	12	188

Table 3: Gadbois' confusion table [33, p.22.5.4].

In order to discriminate among constant amplitude modulation type, Gadbois proposed three additional features: R_2, R_3 and R_4. The main difference between FM, FSK and PSK resides in the phase characteristics. To exploit and extract relevant characteristics, a Digital Phase Lock Loop (DPLL) is used. The output of this PLL, being proportional to the phase derivative, can be roughly summarized as follows:

–for FM, the output is a Gaussian random process

–for FSK, the output is a noisy square wave

–for PSK, the output is uniformly spaced impulses

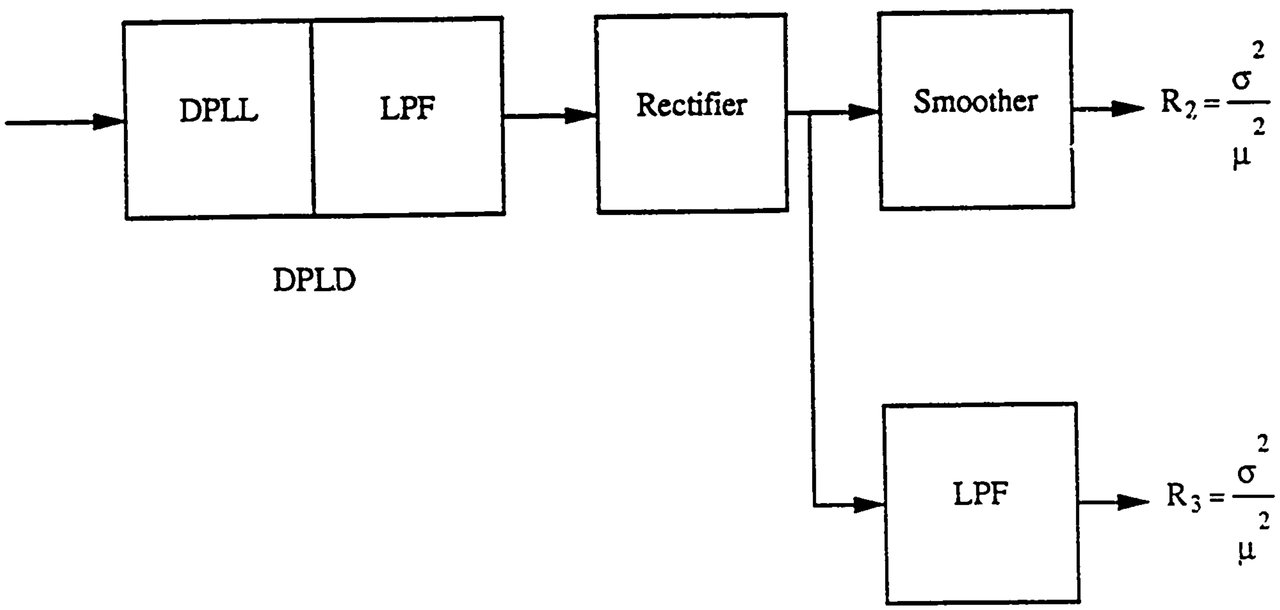

Figure 13: Gadbois' system block diagram [34, p.39]

The discriminative characteristics in the waveforms can be increased by further processing, as shown in Figure 13. The output of the rectifier (DC block rectifier used to remove the carrier) is not significantly modified, except for an FSK signal which generates an almost pure DC signal. Thus, R_2, being a measure of the signal variability, is potentially able to differentiate the three modulation types. However, simulations showed that R_2 alone could hardly separate FM from PSK at SNR less than 20 dB. By adding a low–pass filter, Gadbois created an additional feature, R_3, which would be similar to R_2 except for PSK signals (impulses corresponding to PSK signals contain high frequency harmonics) Therefore, the ratio R_2/R_3, denoted R_4, should provide the separation between FM and PSK:

– for FM, R_3 would be less than R_2 because the noise would be reduced by the filter
– for PSK, R_3 would be much less than R_2 since both noise and some of the baseband signal would be filtered.

The accuracy obtained with these features is 98%, with a SNR of 10 dB, assuming that only these three modulation types are present.

3.2.2 Ribble

Using methods similar to Gadbois, Ribble [9] developed his system around the same features R_1 and R_2, but replaced R_3 and R_4 by ACPOW (see Figure 14).

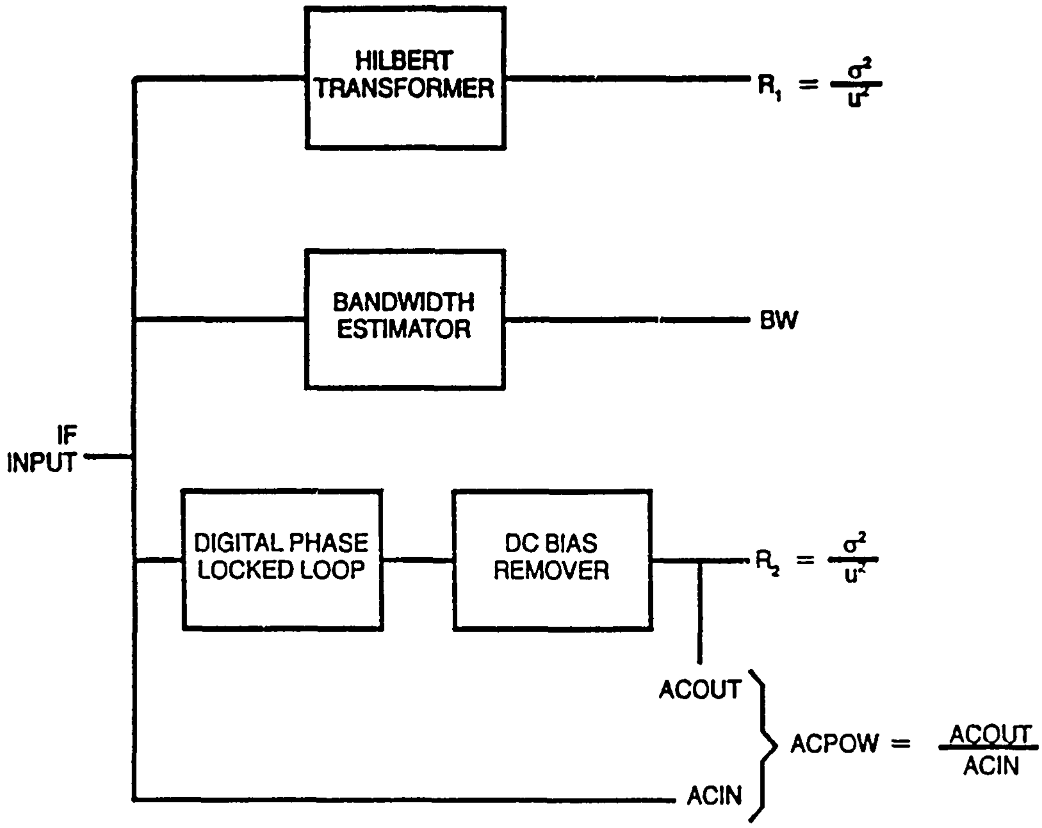

Figure 14: Ribble's system block diagram [9, p.20].

In the block diagram, R_1 is Gadbois' ratio $\hat{R}$, and R_2 is almost like Gadbois'. The last feature is a power parameter, ACPOW. It measures the power at the DPLL output[1]. Thus it gives the value of the amount of phase/frequency changes in the signal. For digital modulation schemes, the value of ACPOW decreases with the baud rate.

Once again the decision process uses a logic tree. The value of $\hat{R}$ is used only to separate Group B (phase modulations) from Group A (amplitude modulations) The other features will complete the estimation of the modulation type. The time required for the

[1]. It is given that ACPOW = ACOUT/ACIN. Here ACIN is a normalisation factor.

computation and the decision is 1.7sec. The performance obtained is about the same as Gadbois', but Ribble reduced the sampling time. He pushed the experiments further and replaced the simulated voice by real voice, obtaining drastically lower accuracies. However Ribble did not go further with the problem and concluded his report with this comment:

> "If the prime modulating signal of concern is voice then a different data acquisition process will be necessary. It would have to reject segments of low/no modulation and stack together those segments which indicate modulation is present. This would obviously result in extremely long time frames to analyze voice, but this seems to be the only way to cope with this situation." Ribble [9, p.72].

3.2.3 UTL

In 1989, DND/DEEM 4 gave a contract to UTL CANADA INC. [36] for an analytical examination of various classification techniques (Jondral, Callaghan, Gadbois, Ribble, Fabrizi, Wakeman and Gardner). The contract also asked for the design of a prototype MRU (Modulation Recognition Unit).

UTL decided that Ribble's approach was the most promising and tried to improve the performance further. The block diagram was roughly the same. However the algorithm for evaluation of the bandwidth was "improved". The logic tree for the decision process was also changed, incorporating a new feature, the product of the preceding features R_1 and R_2. The hardware was set with the following conditions: IF frequency of 34kHz with 12kHz bandwidth and a sampling frequency of 140kHz. The performance was not significantly improved and the results for real voice were still unsatisfying (average probability of success around 25%).

3.2.4 Gallant

Neither Ribble nor UTL were able to obtain good performance with Gadbois' ratio. Gallant [3] looked at the problem again and got very interesting results.

First of all he considered the problem of real voice signals. The preceding authors got good results with simulated voice, but their performance fell drastically with real voice. They all agreed that the gaps in voice signals create "unmodulated" segments which are improper to classify the signal.

> "1. Unmodulated AM and FM produce a pure carrier signal; hence they are indistinguishable from each other and CW; and
>
> 2. Unmodulated SSB and DSB produce no transmitted signal; hence they are indistinguishable from each other and noise" Gallant [3, p.24]

In order to solve this problem, Gallant added a preprocessor to Gadbois' algorithm. before computing the ratio $\hat{R}$, the preprocessor removes the unmodulated segments. This preprocessor is comprised of three steps which will now be described.

The first two steps permit the MRD to reject long stretches of unmodulated waveform on a segment to segment basis[1]. This procedure is performed by the *Front–End* which is divided in two parts: the first part rejects low–energy segments, the second rejects noise–like segments.

Looking at the behavior of $\hat{R}$, Gallant found that "anomalies" occur when the modulated signal envelope is small. The third step consists of rejecting low–valued envelope sampling points, on a point to point basis. When N good points are collected, a "pseudo fully–modulated" segment of 2048 points is formed.

The improvement obtained by this preprocessing prior to the computation of $\hat{R}$ is quite significant. Figures 15 and 16 show the system block diagram and the results with and without the preprocessing.

As we can see in Figure 16, there is no real difference between AM and FM, especially for low modulation index AM. Therefore Gallant proposed a new feature, denoted VAR, able to discriminate very low modulation index AM[2]. This new feature is the variance of the segment to segment envelope variance.

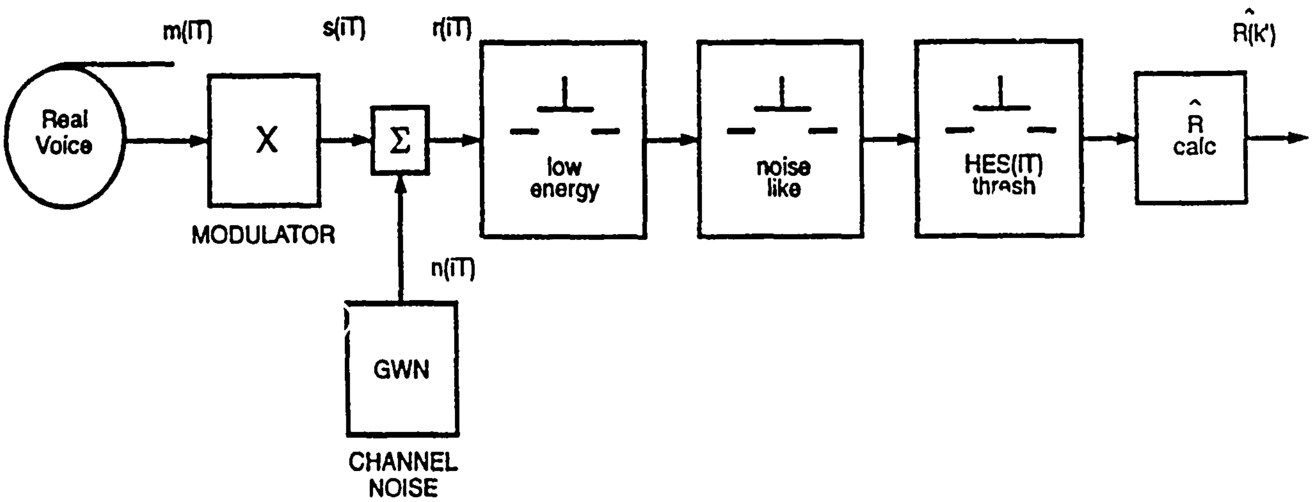

Figure 15: Experiment block diagram [3, p.47].

[1] A segment is formed by N=2048 points and takes 64ms at a sampling rate of 32kHz The complete classification process requires L_2 segments.

[2]. "...some military radios are known to have modulation indices in the range 45% to 65%..." Gallant [3, p.49].

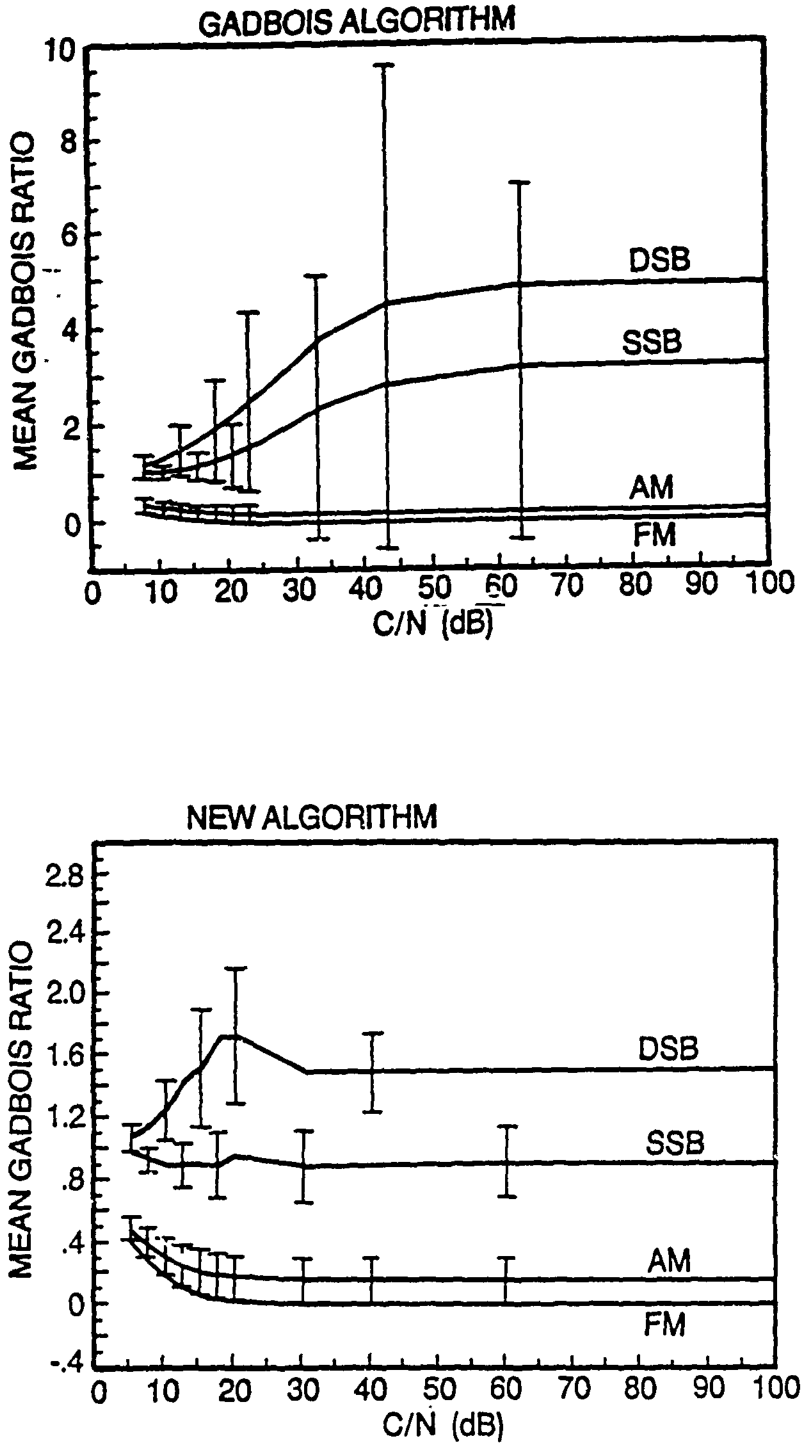

Figure 16: Variance of $\hat{R}$ diminished by preprocessing.

a) Gadbois' ratio $\hat{R}$ (Gallant's experimental results) [3, p.23]
b) New feature proposed by Gallant [3, p.48]

The new feature VAR is computed as followed:

$$\text{MEAN} = E\{\hat{\sigma}^2(k)\} = \frac{1}{L_2}\sum_{k=1}^{L_2}\hat{\sigma}^2(k)$$

$$\text{VAR}\{\hat{\sigma}^2(k)\} = \frac{1}{L_2}\sum_{k=1}^{L_2}[\hat{\sigma}^2(k) - \text{MEAN}]^2$$

L_2: The number of segments which passed the front–end section.

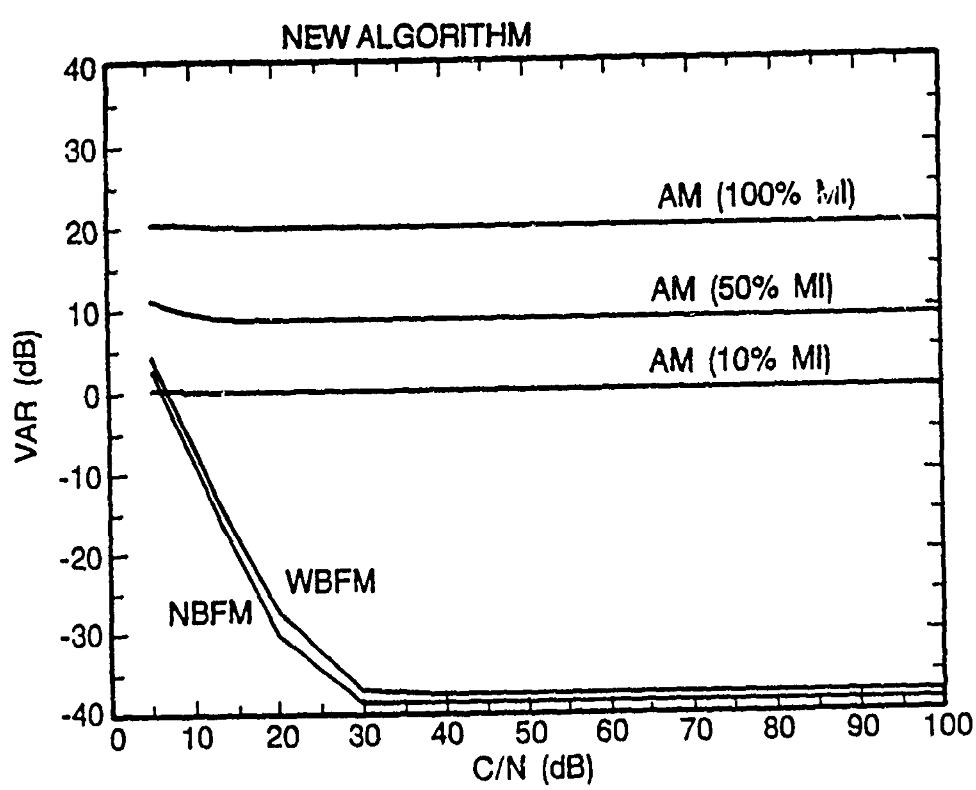

Figure 17: The new feature VAR proposed by Gallant [3, p.54].

The decision process is also a logic tree, however much simpler than Ribble's because there are only two features. The performance obtained with real voice and SNRs above 8dB is quite good (90%). However the acquisition time is long : at least 1.5 seconds.

3.3 OTHERS

3.3.1 Weaver

Probably the first author to publish about modulation type classification is Weaver [37] in 1969. He proposed the use of pattern–recognition techniques to automatically identify the type of modulation on HF radio signals. The intercepted signal was passed through an 8kHz analog bandpass filter and down–converted to baseband. It was then digitized with an 8–bit ADC at a 16kHz sampling rate. The resulting digital signal fed a bank of 29 parallel narrowband filters and envelope detectors. The outputs were averaged over an observation time of one second. These mean values created a 29–dimension feature vector used for the decision process. The classifier is an analog implementation of a linear classifier. Weaver claims 95% classification accuracy at "typical" SNRs for separating AM and SSB. He warns that speech breaks give rise to unmodulated signals and a real system may require several seconds of observation time.

3.3.2 Winkler

Winkler [22][1] proposed a technique based on characterizing the amplitude and phase spectrum. He used a pure sinusoid as the modulating signal. He obtained 100% accuracy, although he admitted that gross classification errors occur when the baseband signal has a noisy spectrum.

3.3.3 Callaghan

Callaghan [38] also based his classifier on the classical pattern–recognition theory, he used a linear classifier. He sampled the signal envelope and amplitude zero–crossing (to get the instantaneous frequency). Then he computed the means and the standard deviations. The sampling frequency was 100Hz and the observation time, 2 seconds. He reported only the performance for separating AM to FM: 99% for an SNR above 20dB. Callaghan concluded his paper admitting that his features were very sensitive to noise.

3.3.4 Fabrizi

Instead of considering the variability of the envelope, Fabrizi [39] suggested a new feature, the envelope peak to mean ratio noted P_e. He also used the mean M of the instantaneous frequency.

---

[1] The information comes from Gallant[gallant, p.9–10], the reference is not available at this time.

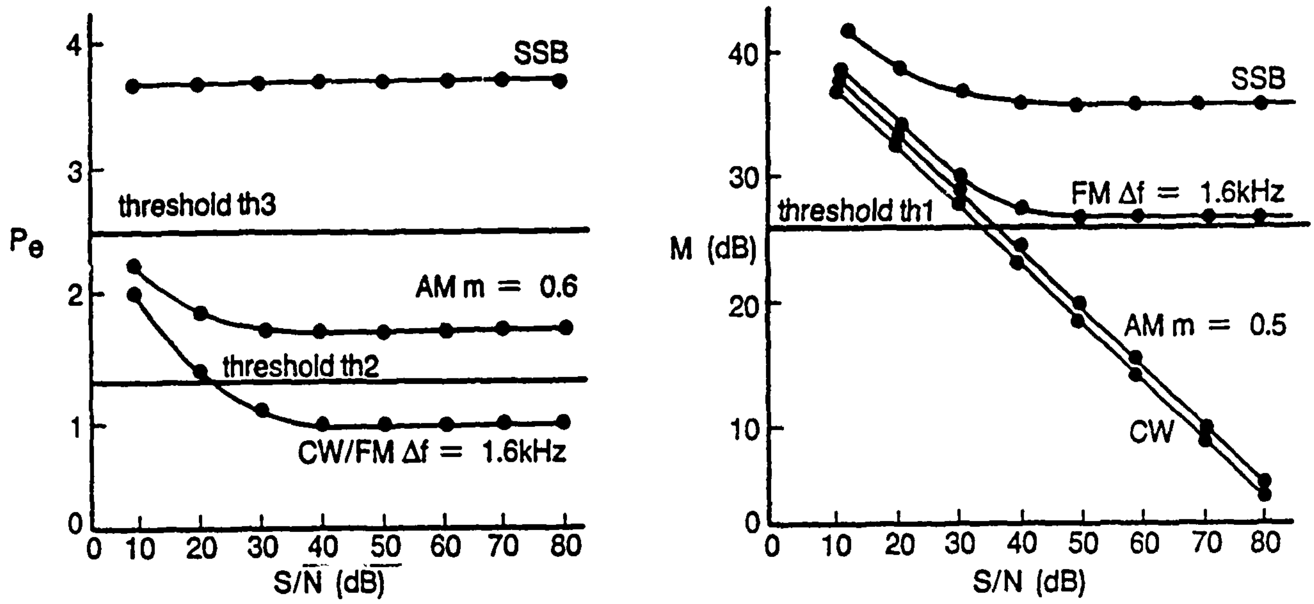

Figure 18: Fabrizi's features [39, p.138].

With only these two features Fabrizi built a logic tree for the decision process (see Figure 19). The parameters were collected over a 250ms sampling time, using 32kHz sampling frequency and 3kHz filtered real voice. Simulations showed that separation of AM from FM using the instantaneous frequency parameter could not be achieved at SNRs below 35dB. However, SSB could be separated from AM/FM group at SNR above 5dB.

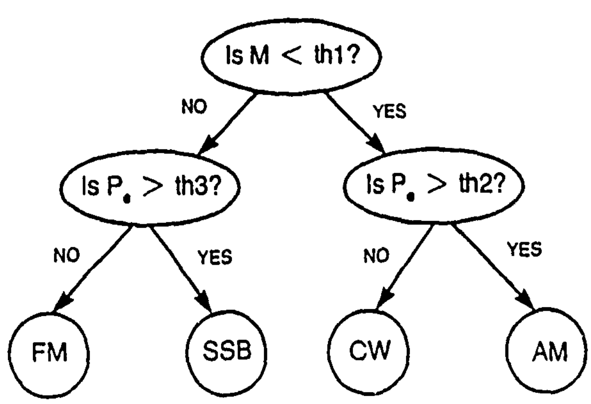

Figure 19: Fabrizi's Decision Tree [39, p.139].

3.3.5 Petrovic

Petrovic [40] proposed a six feature classification algorithm. The basic configuration is shown in Figure 20.

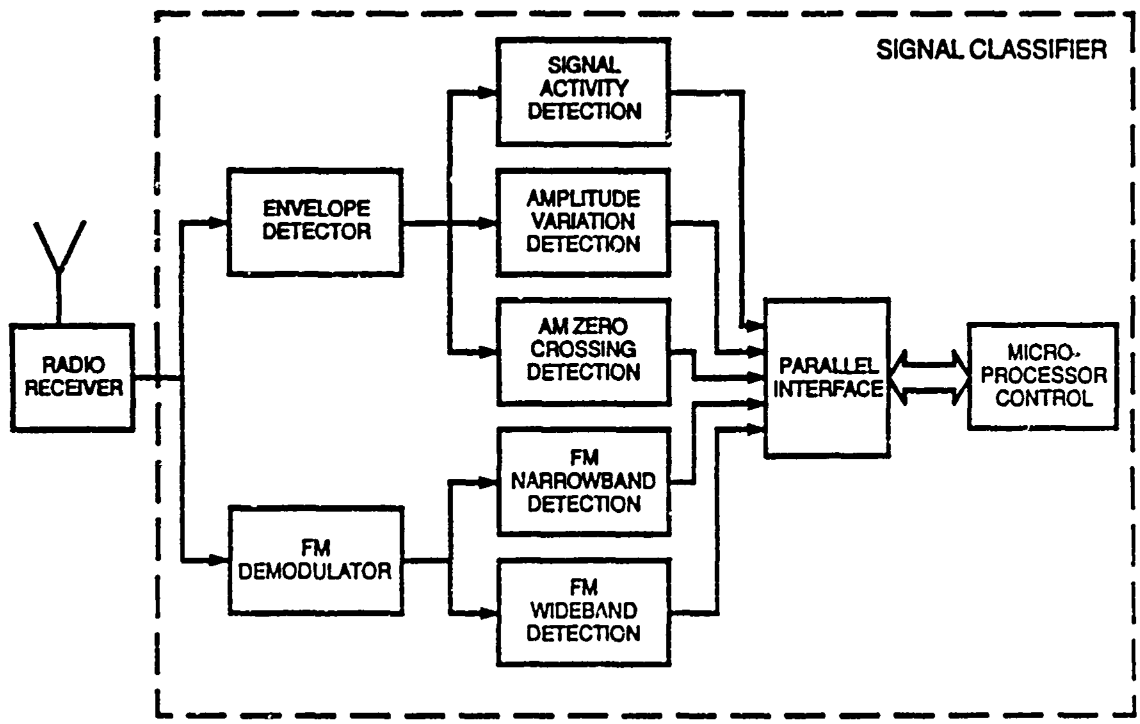

Figure 20: Petrovic's classifier [40, p.387].

The signal activity feature of Petrovic's classifier enables it to determine whether or not a signal is present. The second activity factor represents the percentage of time the envelope exceeds a predetermined threshold. The amplitude variation feature is the mean value of the full–wave rectified AC signal envelope. The FM demodulator gives the instantaneous frequency of the signal. The signal is applied through a low–pass filter (3kHz) and the output is averaged to detect narrowband (analog) FM. The signal is also applied to a 3kHz high–pass filtered to detect wideband (digital) FM.

The decision process is accomplished with a reference logic table (variation of a decision tree). The IF frequency is not given but the author said that the IF bandwidth is optimal at 15kHz. The Confusion Matrix is not given.

3.4 AISBETT APPROACH

A lot of work has been done at the Electronic Research Laboratory, Department of Defence, Defence Research Centre Salisbury, Adelaide, South Australia regarding modulation recognition. New features , which are claimed to be "noise resistant" are presented by two authors, Aisbett and Einicke in this section.

3.4.1 Aisbett

Almost every paper presented proposed features which are very sensitive to noise. For example, remember that Fabrizi was not able to separate AM from FM at SNRs below 35 dB. Aisbett [41][42] closely considered the problem of additive white Gaussian noise (AWGN) regarding time–invariant features for pattern recognition.

She observed that most published modulation recognition schemes perform poorly because the authors chose signal parameters which can only be estimated with a bias in the presence of AWGN. She shows, for example, how the sample mean and sample variance of the instantaneous frequency varies with SNR for a number of analog modulation types Thus she proposed time–domain signal parameters which are unbiased estimators of the true signal parameters in the presence of AWGN with symmetric spectral density.

The three new proposed noise resistant features are A^2, AA' and $A^2\theta'$, where A is the signal envelope, A' the signal envelope derivative and θ' the instantaneous frequency Considering very low SNRs (3dB), she claims discrimination between AM, FM DSB and CW possible on the basis of characterizing the new parameters' statistical distribution functions.

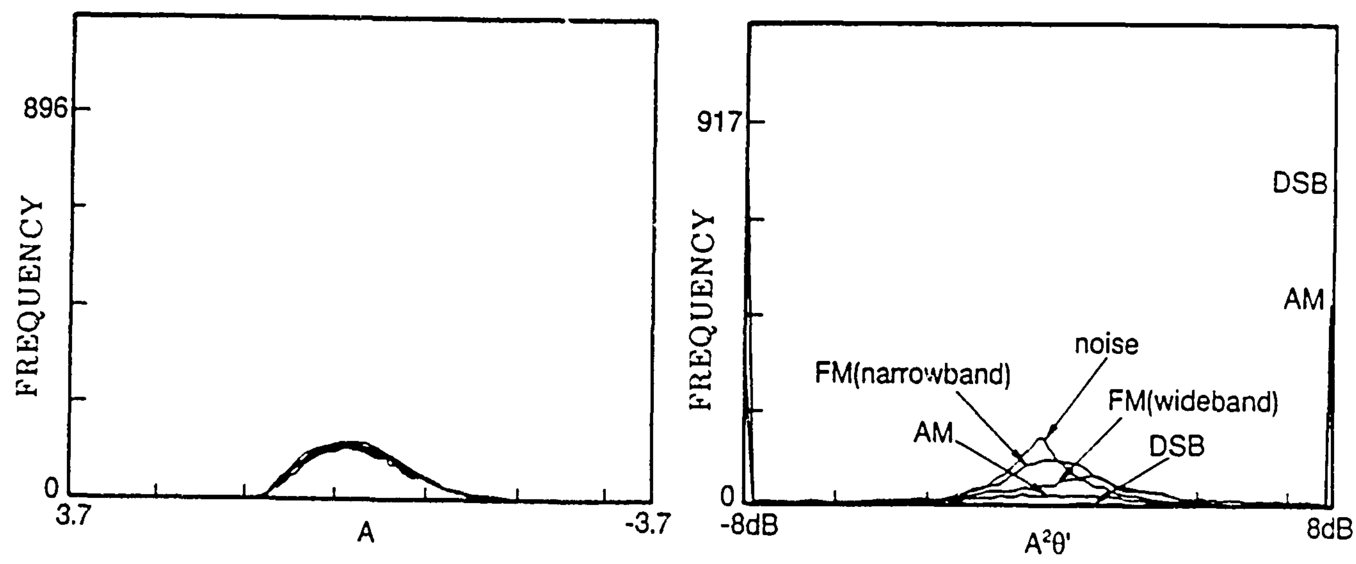

Figure 21: Distributions of Aisbett's parameter $A^2\theta'$
and the typical parameter A (SNR = 1dB) [41, p.12, 17].

It is shown in Figure 21 that for very low SNR (–1 to 3dB), the typical parameter A (signal envelope) has a distribution function which offers no discrimination between the modulation types. However the distribution functions for Aisbett's parameter $A^2\theta'$ are distinct and could therefore be used for classification.

She was satisfied by her preliminary work and envisaged further simulations with a much larger data base in order to implement a classifier using a pattern–recognition algorithm applied to these new features.

3.4.2 Einicke

A few years later, a paper written by Einicke [6] was published. He described a classifier based on Aisbett's features.

$$A^2 = I^2 + Q^2$$
$$AA' = II' + QQ'$$
$$A^2\theta' = IQ' - I'Q$$

He added two classical parameters, the signal envelope and instantaneous frequency

$$A = (A^2)^{1/2}$$
$$F = \theta' = (A^2\theta')/A^2$$

As stated by Aisbett, the statistical distribution of these parameters should have a good discriminating power even for low SNRs. Jondral's method[1] of using the histogram as a features vector is potentially very powerful. However the computation required is too exhaustive for a real time system. One of the best ways to describe a statistical distribution is to use the standard deviation σ, the coefficient of skewness γ and the kurtosis β.

The feature vectors are computed for each modulation type according to three classes of signals: strong signal (30dB), medium (15dB) and weak (5dB). The parameters in the data base are obtained through digital signal processing of data generated from 12–bit ADC having a sampling rate of 20 kHz. For the decision process, a linear classifier (Fisher's functions) is applied to these reference feature vectors. He also tried a quadratic classifier but found no improvement. The performance obtained depends on the sample acquisition time. For an acquisition time of 409ms, the overall performance is around 94% (5dB < SNR < 30dB).

[1] See the digital modulation section.

3.5 HIPP APPROACH

3.5.1 Hipp

This area of modulation classification is very intuitive. The choice of the features depends upon the author's background, imagination, etc. Hipp [8] used a different way to select his features. He did a systematic and exhaustive evaluation of more than twenty features from which he retained only six. The results obtained are very impressive: he classified almost all modulation types (digital and analog) with an overall performance of 95% with SNRs going down to 10dB. His paper is presented in a distinct section for these reasons.

The features used are based on statistical central moments: standard deviation, standard deviation divided by its mean, skewness and kurtosis. These statistical parameters are used to describe the following signal characteristics: amplitude, phase, frequency, spectrum, squared signal spectrum, in–phase channel and quadrature channel

	Amp	Phase	Cos Phase	Sin Phase	Freq	Spectrum[i]	Squared Spectrum[ii]
Standard Deviation	S	X	X	X	X	S	S
Skewness	S	X	X	X	X	X	X
Kurtosis	X	X	X	X	X	X	X
Phase Vector Quality Factor $\sqrt{\cos^2\phi+\sin^2\phi}$							S
RMS Amplitude							X
Signal Standard Deviation of Spectrum[ii]							S

<u>Notes</u>:
S: Hipp's selected features
X: Not selected features
i: Threshold at 3 dB above noise floor
ii: Threshold at 3 dB below spectral peak

Table 4: Features evaluated by Hipp [8, p.20.2.5]

To evaluate all of these features, Hipp collected 400 sets of 1024 data points for each modulation type with SNRs varying from 100dB to 10dB. Then he made a statistical data analysis on the data base in order to select those parameters having greatest discriminating properties. He achieved that evaluation with a stepwise discriminant analysis on a linear classifier (Fisher's functions).

The resulting six–feature vectors allowed the classification of an unknown signal with an overall probability of 95%. The features retained are: the amplitude standard deviation, amplitude skewness, phase spread, spectrum standard deviation (threshold at 3dB above noise floor), spectrum standard deviation (threshold at 3dB below peak) and squared signal spectrum standard deviation (threshold at 3dB below peak).

The simulation was done with a sample acquisition time of 26ms at a sampling frequency of 40kHz for an IF of 10kHz. The carrier frequency was randomly selected within 1kHz of the nominal frequency. The IF bandwidth was fixed at 20kHz. Unfortunately, some modulation parameters remained constant during the simulations: AM modulation index fixed at 90%, FSK frequency deviation 500Hz, baud rate 300 and 1200Baud, and FM frequency deviation at 3 and 5kHz. Moreover, it is not specified how the modulating signal (for analog modulation schemes) was generated. As shown by Ribble, UTL, Gallant, and Fry, simulated voice and real voice do not perform in the same way.

4.0 MR FOR DIGITAL MODULATIONS

The interest in digital modulation classification is growing yet the number of publications is still small. For this reason all the classifiers are represented by only two groups. The first one is represented by Liedtke and Jondral. These two authors, especially the latter, are very well known and are cited in almost every paper on the topic. The second group includes authors who used a different approach to Liedtke's.

4.1 LIEDTKE AND JONDRAL APPROACH

4.1.1 Liedtke

One of the first authors to publish about modulation type identification, Liedtke [2] was also the first to present the concept of modulation recognition applied to digital modulation schemes. The system proposed by the author is fully digital, as shown in Figure 22. The output of the receiver is digitized (In–phase and Quadrature channels) and then filtered by a bank of parallel FIR narrow–band filters. These filters have the same center frequency but different bandwidths. "The best classification result will be automatically obtained behind that filter which matches the signal bandwidth best."

The feature extraction is accomplished with a "universal demodulator". These features are the amplitude, phase and instantaneous frequency (see Figure 23). To get the parameter values in a synchronous way, a timing recovery procedure is used. The parameters are defined as:

$$\text{Amplitude} = (I^2 + Q^2)^{1/2}$$

$$\text{Instantaneous Frequency} = \frac{1}{2\pi}\left[\frac{d\varphi(t)}{dt}\right]_{t=t_i}$$

$$\approx \frac{1}{2\pi}\,\frac{\varphi_{i+1} - \varphi_{i-1}}{2\,\Delta t}$$

$$\varphi = \arctan\,(Q/I)$$

The values are compiled to form histograms of amplitude, frequency and phase. These histograms are used as features for the classification. The 256–point phase histogram of BPSK has peaks at 0 and 180°. For QPSK, the peaks are located at 0, 90, 180 and 270°. And likewise for 8–PSK. Therefore the phase histogram is used to classify these modulation types. The object is to use the histogram as input to a procedure that will recognize the number of phases. A pattern–recognition algorithm could be used, taking each cell of the histogram as an element of the feature vector. However, to simplify the computation, Liedtke used suboptimal weighting functions, as shown in Figure 24

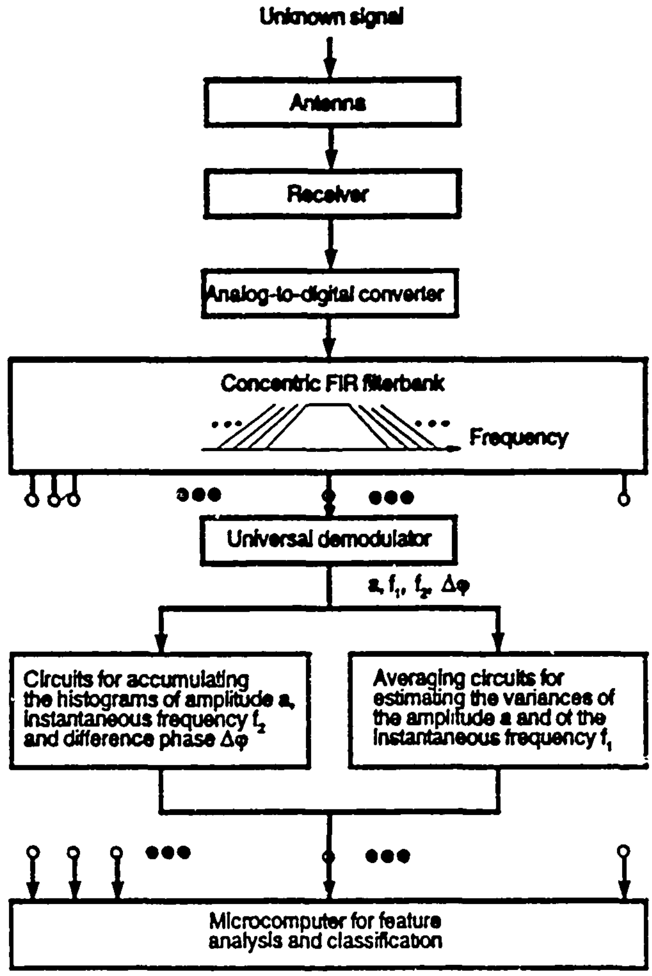

Figure 22: Block diagram of Liedtke's classifier [2, p.313].

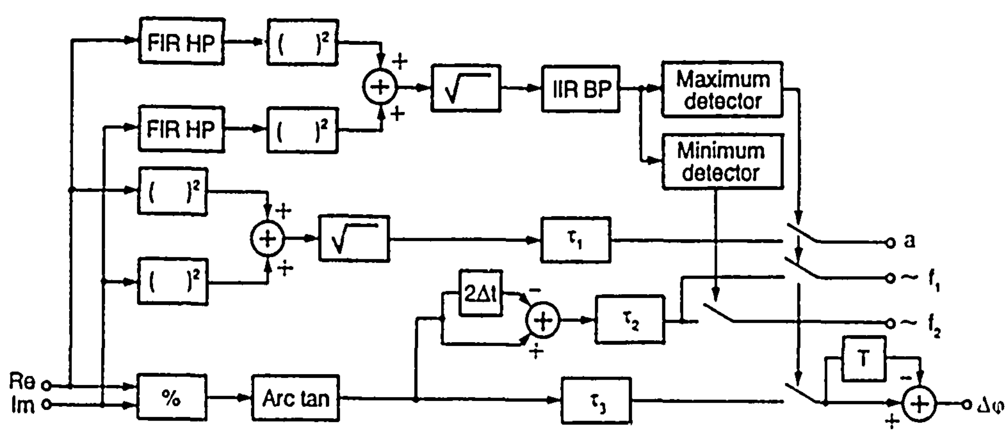

Figure 23: Liedtke's universal demodulator [2, p.314].

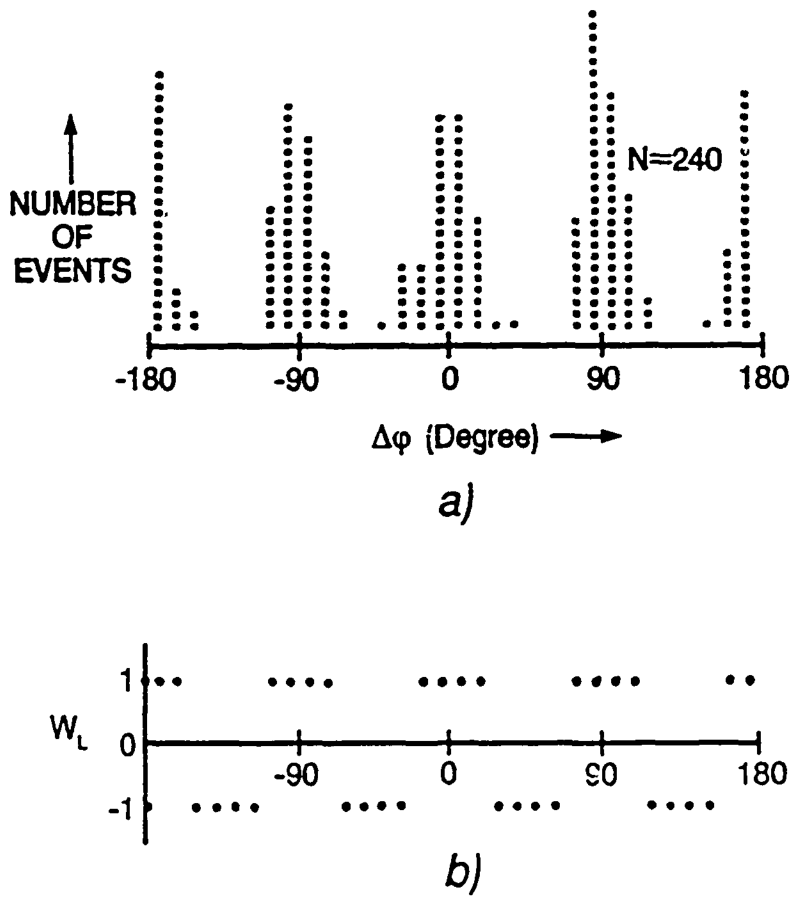

Figure 24: Classification of PSK signals by Liedtke
a) Actual phase histogram for QPSK [2, p.315]
b) Suboptimal weighting function for QPSK [2, p.316]

The unknown signal phase histogram is compared to the weighting functions and a value of similarity is attributed to each. The largest output is retained and noted DPHI, corresponding to the number of phases detected (2, 4 or 8).

If the recognition criterion of PSK modulation types is not satisfied, the classification of ASK and FSK is investigated. To identify these modulations, the variances of the amplitude and frequency (AVAR and FVAR respectively) are calculated. A large AVAR indicates ASK, and a large FVAR, FSK.

In a similar way to the phase histogram, the amplitude and frequency histograms are also used. The amplitude histogram of ASK contains two peaks, as does the frequency histogram for FSK. These histograms should contain only one peak for other modulation schemes. The resulting variables are respectively AHI and FHI.

The decision process is accomplished with three boolean equations: one for the PSK modulation type, one for ASK and one for FSK. The decision function for PSK is

$$[\,(\max_{w_i}(\mathrm{DPHI}))_{i\geq 2} > \mathrm{TDPHI}]\cdot[\mathrm{AVAR} < \mathrm{TLAVAR}]\cdot[\mathrm{FVAR} < \mathrm{TFVAR}] = \mathrm{TRUE}$$

i=2,4 or 8 is the number of phases
TDPHI is the threshold for the phase histogram
TLAVAR is the threshold for the amplitude variance
TFVAR is the threshold for the frequency variance
the symbol · is a logical AND

If the function is "true" for i=2, BPSK is detected; if i=4, QPSK is detected; and if i=8, 8–PSK is present. Otherwise, i.e. DPHI is optimal for i=1, the test continues for ASK and FSK. The equations are the followings.

$$[\mathrm{AHI} > \mathrm{TAHI}]\cdot[\mathrm{AVAR} > \mathrm{TUAVAR}] = \mathrm{TRUE}\ldots\text{for ASK and}$$
$$[\mathrm{FHI} > \mathrm{TFHI}]\cdot[\mathrm{FVAR} > \mathrm{TFVAR}]\cdot[\mathrm{AVAR} < \mathrm{TLAVAR}] = \mathrm{TRUE}\ldots\text{for FSK}$$

TAHI is the threshold for AHI
TUAVAR is the upper threshold of amplitude variance
TFHI is the threshold for FHI

These functions and the thresholds are schematized in Figure 25. The five separation parameters are shown together. The dashed lines point out which classes are separated by which separation parameters. The overlapping between PSK and FSK indicates some remaining difficulties.

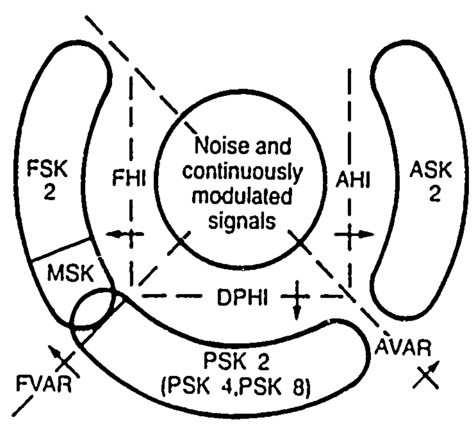

Figure 25: Schematized class space with separation parameters [2, p.318].

The results shown in Figure 26 indicate a classifier very sensitive to noise.

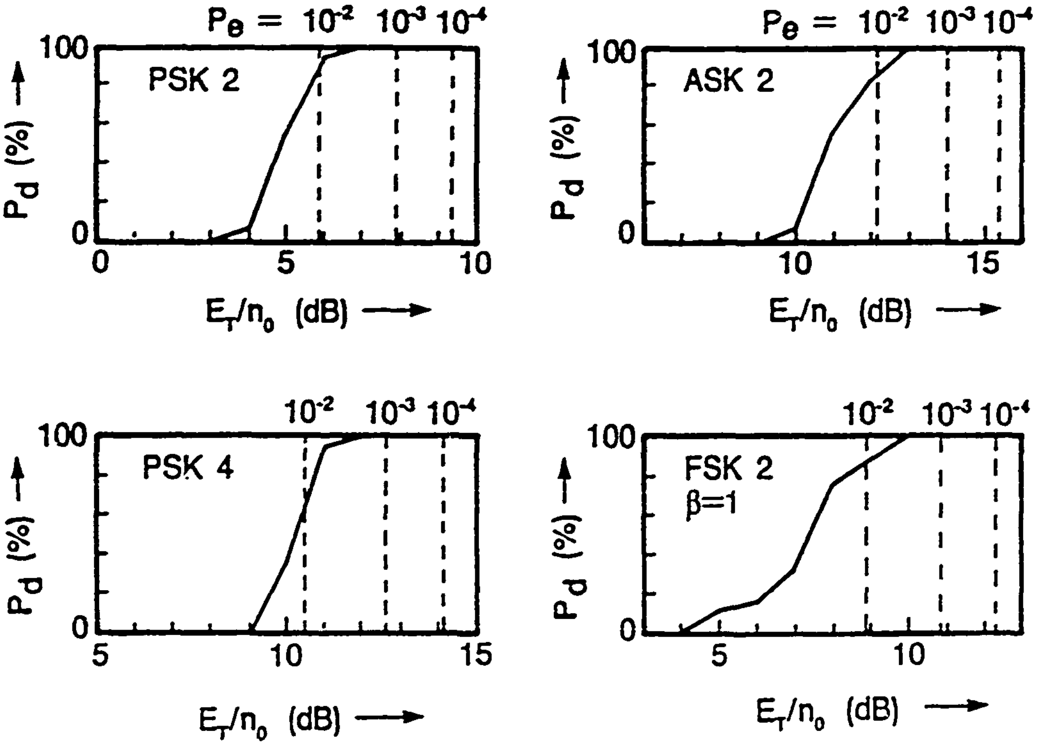

Figure 26: Liedtke's results [2, p.319].

4.1.2 Jondral

Jondral [28][29] used an approach similar to Liedtke's to which he added two analog modulation types (AM and SSB) and pattern–recognition techniques for the decision process.

The features used by the author are based on histograms. However, he did not use Liedtke's synchronization system. Therefore the histograms are different, although the parameters are similar: amplitude, instantaneous frequency and phase. The phase characteristic called *zero phase*, used to detect BPSK, is obtained by squaring the signal to create a carrier at twice the frequency which can be caught by a tracking loop, resulting in the detection of 0 and π phases in BPSK signal.

Collecting the parameters for 4096 points, a histogram of 192 cells is computed (see Figure 27). This histogram is used as the feature vector. However, Jondral realized that the features between positions 60 and 140 do not contribute to the discrimination. The final results were obtained with 93–D feature vectors. This is remarkably bigger than Gadbois, who used only a few features.

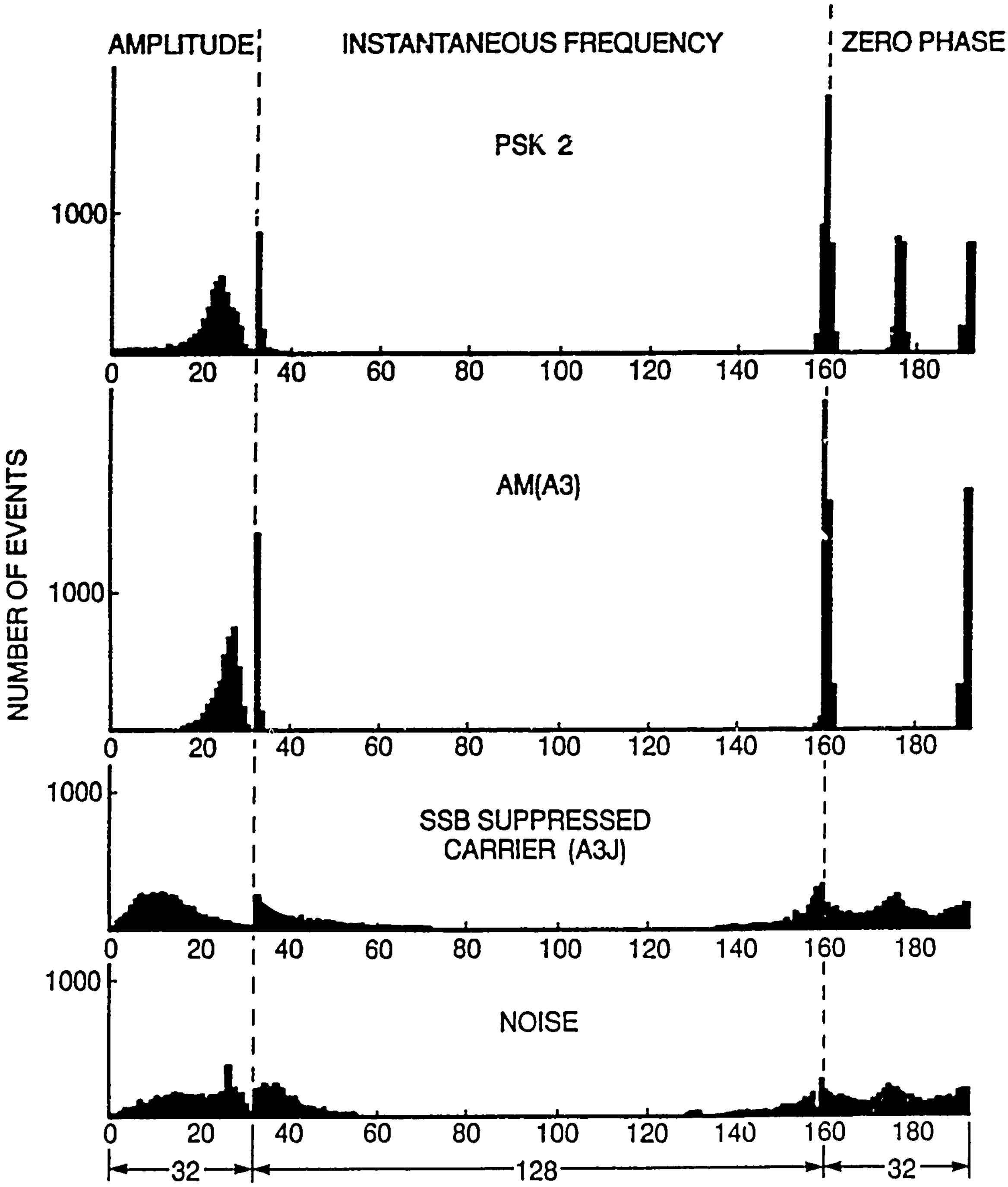

Figure 27: Examples of histograms [28, p.185].

The performance obtained is very good, but the SNRs are not specified. The system operated for a signal down–converted at 200kHz. The 12–bit ADC sampling frequency is chosen according to the IF signal bandwidth: 2.2kHz to 22.2kHz respectively for 300Hz to 6kHz. To acquire the 4096 data points, an acquisition time of 0.18 to 1.8 seconds was required. The decision function was based on a pattern–recognition technique. Jondral tried both linear and quadratic classifiers (minimum mean squared error): he respectively got an overall accuracy of 93% and 98%. To reduce the number of terms for the quadratic classifier, he applied the Karhunen–Loève–transform to the feature vectors. That way, 30 transformed–features (containing about 97% of the information) were used in the classifier instead of 93 features.

4.1.3 Dominguez

Also using a histogram as the feature vector, Dominguez [43][44] realized a recognition system very similar to Jondral's. The parameters used in his histogram are the amplitude, the instantaneous frequency and the instantaneous phase $\varphi(k)$.

$$\varphi(k) = \arctan\{Q(k)/I(k)\}$$
$$k < 3000 \text{ points}$$

The dimension of the vectors is 79: 31 components corresponding to the amplitude; 31 components for the frequency; and 17 for the phase. The classifier was linear. The overall performance is 95%, and the system recognizes all modulation types. However, the SNR is not given, neither is the message signal used for analog schemes.

Dominguez also presented in [43] a preprocessor. Noticing that the system works better if the signal is perfectly centered over the IF frequency, he proposed a preprocessing able to estimate the frequency of a carrier. This preprocessor also explores the spectrum to detect adjacent modulated signals: it is similar to an energy detection subsystem.

To analyze the spectrum, a periodogram is used. First, the spectrum is differentiated to detect carriers. Then the symmetry around the carriers is studied. The output is the number of signals. The classification algorithm will be processed if there is only one signal. An example is shown in Figure 28. The symmetry around the first carrier indicates a first modulation type spread on both sides of the carrier A second carrier is present in the spectrum, indicating a second modulation type. This signal would have to be filtered by the preprocessor to allow the first signal to be recognized.

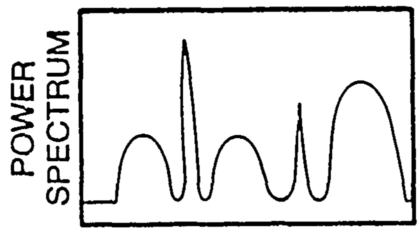

Figure 28: Intercepted spectrum[43]

4.1.4 Adams

Adams [45] proposed an improvement to Jondral's recognition system by using a new classification process. On the same 192 features, he applied the PCA (Principal Component Analysis) algorithm to reduce the dimension of the pattern vector. The author did not give the size of the resulting vector. Then the MANOVA algorithm is applied on these new vectors to optimize the discrimination. This is a linear discrimination technique. MANOVA is a conventional multivariate statistical technique discussed respectively in [46].

No performance results are given in the paper. It is not obvious that this technique would perform better than Jondral's. Jondral also used data reduction, reducing the dimension of his vectors from 192 to 30 with the Karhunen–Loève–transform, keeping 97% of the information. Moreover Jondral used a multivariate linear and quadratic discrimination techniques.

4.2 OTHERS

4.2.1 Mammone

The paper proposed by Mammone [10] presents a recognition system for PSK signals. Therefore only two modulation schemes are concerned. However, he also presented a technique for evaluation of the bit rate. The phase derivative is used to find the transitions which occur between every data symbol.

The received signal is digitized and expressed as z (k). The phase p (k) and its derivative are expressed by:

$$p\,(k) = \arctan\{\mathrm{Im}[z\,(k)]/\mathrm{Re}[z\,(k)]\}$$
$$p'\,(k) = p\,(k) - p\,(k-1)$$

The carrier frequency is found by averaging p' (average of the instantaneous frequency). The bit rate estimation requires further processing. The intervals between pulses are multiples of the symbol rate, which is, knowing the modulation type, indicating the bit rate. The amplitude of the pulses indicates the phase shift. $\pi/2$ or π. Therefore it is a feature to discriminate BPSK from QPSK. However, because the signals are filtered and phase noise is present, the instantaneous phase shifts are not very accurately representing the true phase shifts. The author presented another approach able to estimate more accurately the amplitude of these phase shifts.

The decision process is accomplished with a logic tree. The performance is given for a system operating with 1024 sampling points, 3kHz and 30kHz bandwidth depending on the baud rate (from 75 to 19200 bps), and C/N_0 between 35 and 70dB. The performance is given according to the baud rate. It is shown that a low baud rate signal (below 200bps) would require a much longer acquisition time (see Table 5).

MOD. TYPE	C/No	DATA RATE	CLASSIFIED AS		
			BPSK	QPSK	CW
	dB	bps	%	%	%
BPSK	45	75 300 1200	82 98 95	15 2 5	3 0 0
QPSK	45	2400 4800 19200	1 16 89	99 83 11	0 1 0
BPSK	60	75 300 1200	82 100 96	14 0 4	4 0 0
QPSK	60	2400 4800 19200	0 2 0	100 98 100	0 0 0

Table 5: Mammone's results [10, p.28.4.6].

4.2.2 DeSimio

For his Master's Degree, DeSimio [4][5] made some simulations concerning ASK, FSK, BPSK and QPSK automatic recognition. Nine features are used. The mean and variance of the signal envelope are used to discriminate ASK signals. The other features are taken from spectra.

The following four features are the spectral location and magnitude of the two largest correlation peaks of the signal spectrum with a sinc^2 reference function. Two large peaks indicate FSK. The results of correlations with spectra of the signal squared and quadrupled provide information related to the number of phases: BPSK vs QPSK

Unfortunately the simulation was done with fixed baud rate signals (2500 bauds), therefore the features presented could not be applied to the general case of unknown bit rate.

The decision process is done with an adaptive technique. The LMS algorithm, which is derived from the perceptron, optimizes the values of weight vectors during a long learning process. Once the learning process is finished, the system is ready for testing with unknown signals. DeSimio tested the classifier with only 16 samples (SNRs from 20 to 5dB) for which he obtained no misclassification.

The perceptron classifier used by DeSimio could be seen, by extension, as a very simple neural network with no hidden layer, nonrecurrent and a hard limiting transfer function. The processing element of the classifier is shown in Figure 29.

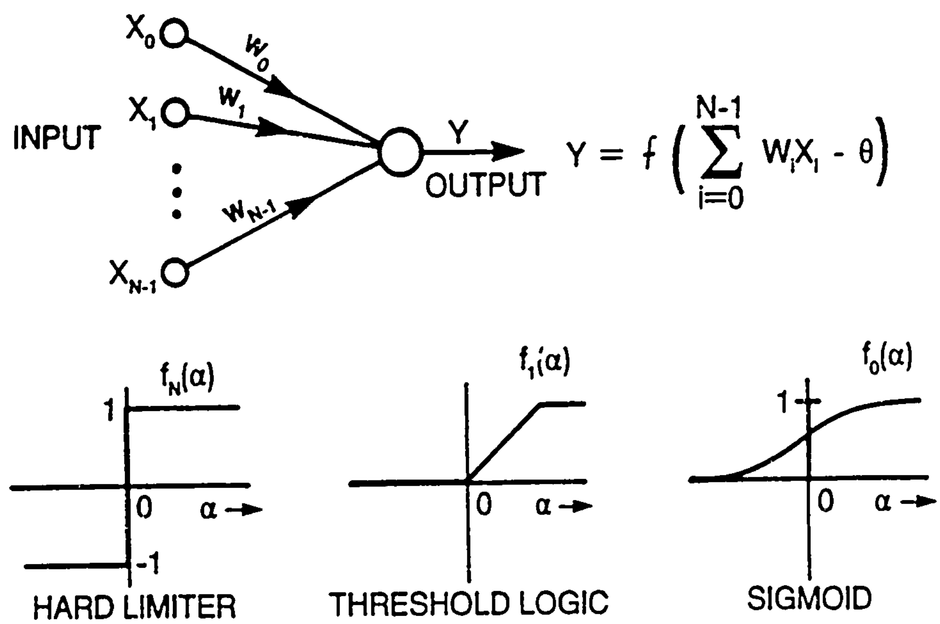

Figure 29: A single unit of the perceptron classifier [47, p.5]

As in the case of the linear classifier, the output is computed by vector multiplications. There are as many outputs in the network as classes, as we had with the linear classifier. However the weighting vectors W are not determined by statistical methods. Via a long learning process (100 000 iterations) the five weight vectors (one per modulation type) are non–parametrically determined. In Figure 29, the transfer function for the perceptron and LMS is represented respectively by *Threshold Logic* and by *Hard Limiter*.

5.0 MR BASED ON ENERGY DETECTION ALGORITHMS

As described so far, modulation recognition has been accomplished using conventional pattern recognition techniques. In this chapter, a different approach is presented, based on the premise that if we are able to detect a modulation type and only one, we recognize it. This chapter presents modulation recognition based on energy detection algorithms.

5.1 READY APPROACH

In 1986, a U.S. Patent was registered for a "Modulation Detector and Classifier" [48]. Ready presented a new approach for some sort of digital modulation classification. It is quite a complex hardware system, so the complete details will not be explained here.

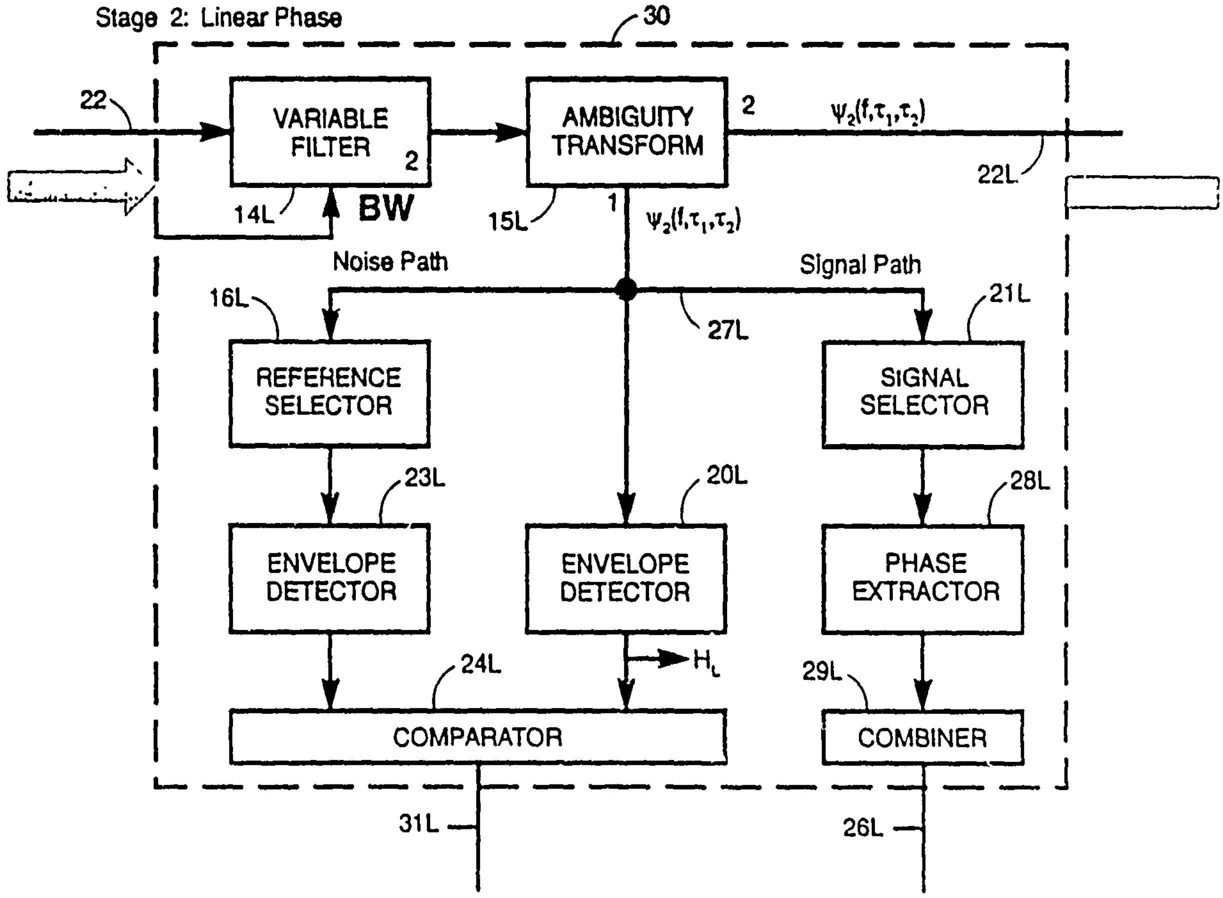

Figure 30: One of the 3 similar stages of Ready's system [48, p 2].

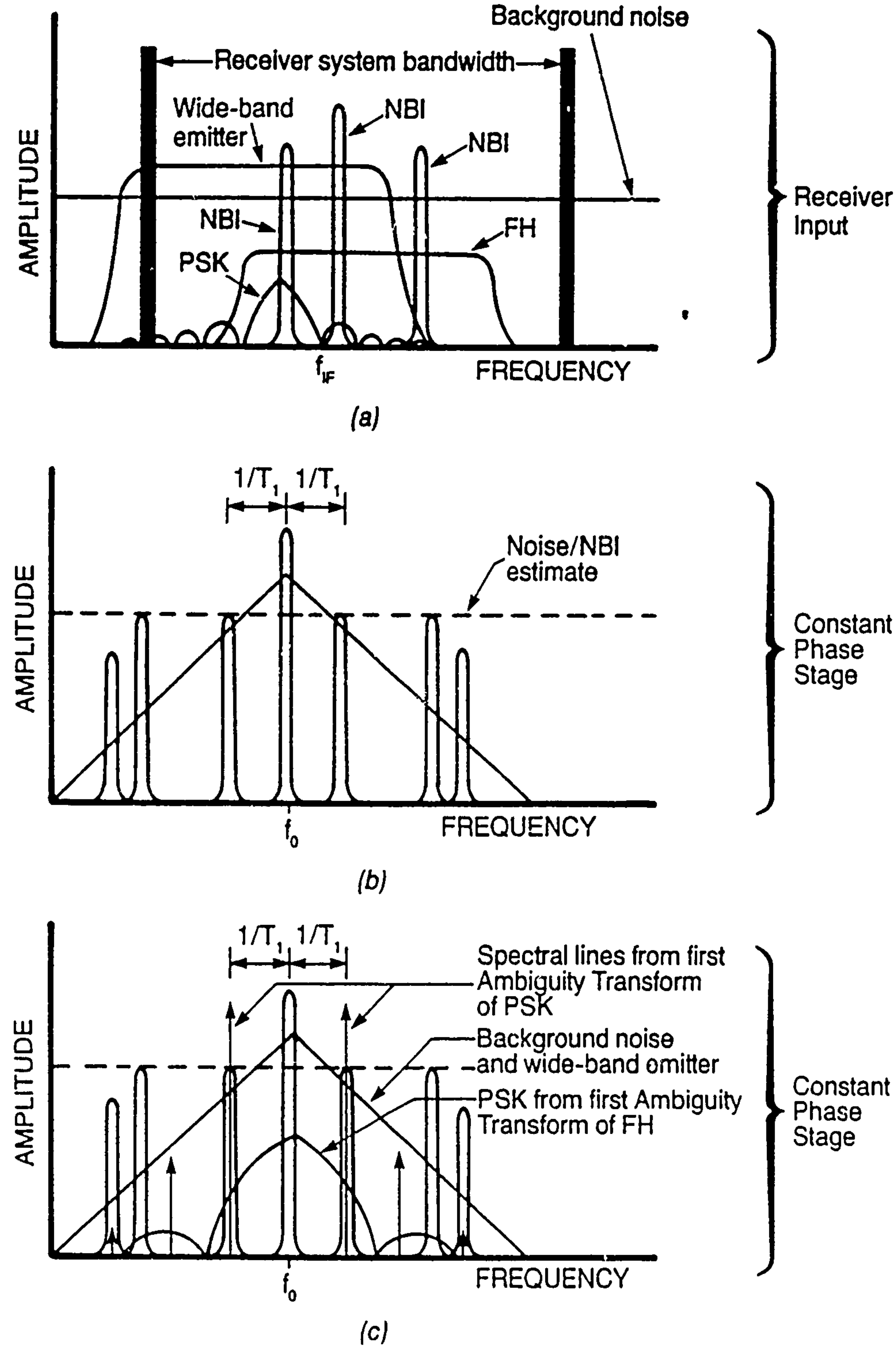

Figure 31: Example of the processing performed by the system
a) Received spectrum b) Noise path c) Signal path [48, p.9–11].

The system is basically made by cascading N similar stages. Stage 1 detects Constant–Phase modulation schemes (M–PSK); stage 2 for Linear–Phase (M–FSK); and stage 3 for Quadratic–Phase signal (linear FM or chirp). Higher order phase modulations could be implemented. The system performs classification by characterizing the phase. This hardware could not be modified for analog modulation types.

In Figure 30, the signal 22 comes from the preceeding stage, and signal 22L goes to the following stage. Signals 31L and 26L are used by the decision circuits. The system performs signal detection and classification. Therefore, it allows more than one modulation type in the bandwidth of interest. The system would also work in the presence of interfering signals.

Figure 31–a shows an example of a received spectrum where several modulation types are present. The detection of a constant–phase signal, in the example PSK, would be performed by stage 1. The output of the first stage noise path envelope detector is an estimation of the noise amplitude. This threshold value is used to detect the presence of a signal at the output of the first stage signal path envelope detector. Figure 31–c, presents that output. We can see the same spectrum as 31–b but with two additional narrow spectral lines above the noise threshold. These spectral lines are symmetric about the carrier frequency and indicate the presence of a constant–phase signal and its symbol rate. The last path, i.e. the phase extractor, gives information on the phase difference between the clock circuit and the data clock of the received signal.

5.2 KIM APPROACH

Kim [49][50] presented his energy detector as a way to classify phase–modulated signals. Using only one feature, i.e. qLLR defined below, he was able to discriminate BPSK from QPSK for very low SNR (0dB). The algorithm is as follows:

$$\mathrm{qLLR} \equiv (\Sigma_I - \Sigma_Q)^2 + 4\Sigma_{IQ}^2$$

where all Σ are energy related and defined by

$$\Sigma_I \equiv \sum_{n=1}^{N} r_{I,n}^2 \; ; \; \Sigma_Q \equiv \sum_{n=1}^{N} r_{Q,n}^2 \; ; \; \Sigma_{IQ} \equiv \sum_{n=1}^{N} r_{I,n}^2 \, r_{Q,n}^2$$

$$r_{I,n} \equiv \int_{(n-1)T_s}^{nT_s} r(t)\cos\omega_C t \; dt; \; n=1,\ldots,N$$

$$r_{Q,n} \equiv \int_{(n-1)T_s}^{nT_s} r(t)\sin\omega_C t \; dt; \; n=1,\ldots,N$$

and r (t) is the received signal containing AWGN.

The feature proposed seems to perform well with AWGN. The results are given for SNRs as low as 0dB with close to 100% correct classification. However, Kim assumed a precise knowledge of the carrier frequency and symbol timing. In a real system, accurate determinations of these might be very difficult for signals as noisy as 0 dB, and might lead to unfortunate performance degradations. Moreover, it is not clear that the system will not react to other modulation schemes that might be presented to the Modulation Recognition Sub–System in a real system. This technique is applicable only to digital phase modulation schemes.

5.3 GARDNER APPROACH

Gardner is well known in the energy detection area. In a paper [51] he proposed the use of his techniques for modulation classification purposes. The interception of LPI signal is accomplished by exploiting cyclic features which are present in modulated signals: sinewave carrier, data rate, etc.. The theory of spectral correlation in cyclostationary signals developed by Gardner has been shown to be a general and flexible cyclic feature estimator called a cyclic spectrum analyser.

The process maps a time varying signal (2–D) into a 3–D distribution, using cyclic spectra. The distribution presents some characteristic shapes potentially useful for modulation classification.

The time varying signal $x(t)$ is cyclostationary if the parameter

$$R_x^{\alpha}(\tau) = \lim_{T\to\infty} \frac{1}{T} \int_{-T/2}^{T/2} x(t+\tau/2)x^*(t-\tau/2)e^{-i2\pi\alpha t}dt$$

is not zero for some $\alpha \neq 0$. $R_x^{\alpha}(\tau)$ is called the cyclic autocorrelation, and its Fourier transform is called the cyclic spectrum, $S_x^{\alpha}(f)$, and can be interpreted as a spectral correlation function via the following characterization.

$$S_x^{\alpha}(f) = \lim_{T\to\infty} \lim_{\Delta t\to\infty} \frac{1}{T\Delta t} \int_{-\Delta t/2}^{\Delta t/2} X_T(t, f+\alpha/2)X_T^*(t, f-\alpha/2)dt$$

$$X_T(t,v) = \int_{t-T/2}^{t+T/2} x(u)e^{-i2\pi vu}du$$

The set $\{\alpha: R_x^{\alpha}(\tau) \neq 0\}$ is called the set of cycle frequencies. Some examples are presented in Figure 32.

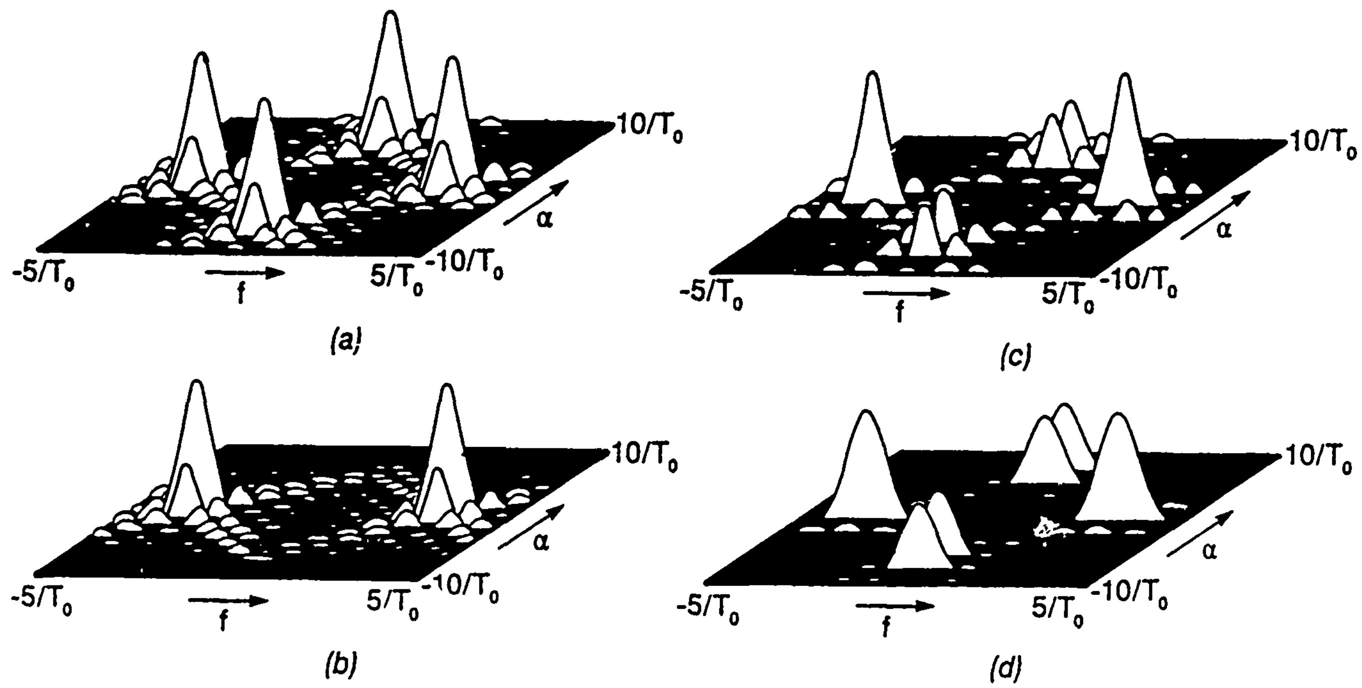

Figure 32: Theorical spectral correlation distributions [52, p.902]
a) BPSK b) QPSK c) OQPSK d) MSK

These distributions show some discriminating power among modulation types. However, Gardner did not discuss how these distributions are sensitive to parameters such as baud rate, modulation index, carrier frequency, etc.. Neither did he discuss the ability of his cyclic feature to characterize analogue modulations regarding the random nature of voice (a nonstationary process). The potential advantage of this technique is its ability to recognize a modulated signal in the presence of interference.

6.0 CONCLUSION

In this document, we have seen how different authors have performed modulation recognition. Using pattern recognition techniques, they have implemented a classification algorithm able to associate features, which have been judged discriminative, to different classes according to the similarity between the unknown pattern vector and the reference pattern vectors obtained during the training.

For the classification process, it seems that the linear and the binary tree classifier have been adopted almost unanimously. The linear classifier is the simplest of parametric techniques. Using only the pooled within–class covariance matrix to classify the unknown signal, this technique is very fast and easily implemented. For a real–time requirement such as MR, it is a good choice. However, for solving bigger problems, with numerous classes and features, it is usually advantageous to use more sophisticated algorithms. The quadratic classifier, by using parabolic functions instead of linear ones to discriminate clusters, could, depending upon the particular application, present improved performance The outcome is an additional amount of memory and computation time for manipulating the within–class covariance matrices (one per class).

Although these parametric techniques could be optimal in some applications (typically when the distributions are normal), the more general case of non–parametric techniques sometimes needs to be investigated (for example, when distributions are multi–modal). In this technical note, only the binary tree concept has been discussed. Modern algorithms for creating trees are very powerful for handling a large amount of classes and features. The CART algorithm is probably the most standard type.

Notice that there also exists other nonparametric algorithms in classical pattern–recognition literature [15]. *Kernel* (Parzen estimator) and *K–nearest–neighbor* approaches are very well known. In the first case, the density function, instead of assumedly being known as in the parametric approach, is approximated by a sum of kernels. These can be any functions, although normal kernels are usually used. For classification, the unknown signal is compared to all N training samples according to the the kernel function in use and is associated to the closest class considering the overall probability. For the K–nearest–neighbor, the decision rule is simplified by considering only the K samples closest to the unknown (instead of all N samples) to compute the probabilities $p(\omega_i | x)$. Unfortunately it is still necessary to store all N samples and compare the unknown with all these samples to find the K closest points. Therefore, the amount of computation is not significantly decreased. In order to overcome this disadvantage, it would be interesting to eliminate samples, keeping only some good representatives (located along boundaries). This process is called *condensed K–nearest–neighbor* decision rule in [15].

Another alternative, not proposed by the authors mentioned in this paper, but very popular in modern pattern recognition research centers, could also be applied to the MR problem. Neural networks are very interesting nonparametric techniques which provide efficiency and extreme versatility with numerous applications. The great interest for this artificial brain is fairly new but some algorithms are becoming well established, such as the back–propagation network. It is highly likely that some new MR systems will use neural networks, in the future, when neural network VLSIs will be available on the market.

Although the choice of the classifier is very important, the selection of good features is fundamental. The features have to be discriminative, noise resistant and requiring as few computation as possible. A feature is obtained from the probability distribution (histogram) of a parameter, which is usually computed on a point to point basis from the sampled signal waveform. Popular parameters are the signal envelope, the instantaneous frequency, etc. Some authors also used the frequency distribution (PSD) to represent the parameter. In both cases, the features are chosen in order to represent efficiently characteristics from the distribution. Very often the mean and standard deviation are used, although the skewness and the kurtosis are also popular (Einicke and Hipp). Another approach consists of using directly the bin values of the distribution as features (Jondral). Unfortunately the latter gives rise to a very large dimension feature vector. To avoid that problem, only the most significant bins can be used: Miller used the first bin, p(0), from the signal envelope distribution. See Table 6 for more examples.

Section	Author	Parameter	Time/Freq	Feature
3.1	Miller	Signal Envelope A	T	Bin value p(0)
3.2	Gadbois	A^2	T	$\frac{\sigma^2(A^2)}{(\overline{A^2})^2}$
	Gallant	A^2	T	$\sigma(\sigma^2(A^2))$
3.3	Fabrizi	Instant. Freq. F_i	T	$\mu(F_i)$
3.4	Aisbett	A^2F_i	T	σ,γ,β
		AA'	T	"
3.5	Hipp	PSD(signal R)	F	$\sigma(PSD(R))$
		PSD(signal^2)	F	$\sigma(PSD(R^2))$
4.1	Jondral	A, F_i and F_i'	T	Bin values

Table 6: Examples of features used for modulation recognition

It would be interesting at this time to investigate a stepwise selection of these features to get a discriminative pattern performing with noisy signals and short acquisition times. Some new features would also be very desirable, features able to deal with lower SNRs, shorter acquisition times, and more modulation types. From a military perspective, classification of signals with SNRs as low as 5 to 10 dB presents a real interest. Gallant's solution of very long acquisition time (a little below 2 seconds) to remove gaps in the voice is not adequate for all applications. A new feature able to discriminate analogue modulation types with a shorter acquisition time would be very desirable. Finally it would be worthwhile to be able to recognize other power–efficient modulation types such as MSK, OQSPK, GMSK and TFM.

REFERENCES

[1] S. Faulkner, "CERES – A Computer–Aided Radio Monitoring and Surveillance System", International Defense Review: An Interavia Publication", June 1977.

[2] Liedtke, F.F., "Computer Simulation of an Automatic Classification Procedure for Digitally Modulated Communication Signals with Unknown Parameters", Signal Processing, Vol.6, 1984.

[3] Gallant, M.J., "Modulation Recognition", Royal Military College, Kingston, ONT., April 1989.

[4] DeSimio, Martin P., "Automatic Classification of Digitally Modulated Signals", Air Force Inst of Technology, Wright–Patterson Air Force Base, Ohio, December 1987.

[5] DeSimio, Martin P. et al., "Adaptive Generation of Decision Functions for Classification of Digitally Modulated Signals", NAECON, 1988.

[6] Einicke,G.A., "A Flexible Architecture for Real–Time Modulation Recognition", Communications Div., Electronics Research Laboratory, Salisbury, S.A.

[7] Fry, Robert, "Modifications to a Spectrum Surveillance Electronic Support Measure (ESM) Receiver Developed for the Department of National Defence (U)", Technical Note 85–7, Defence Research Establishement Ottawa, March 1985.

[8] Hipp, Jackie E., "Modulation Classification Based on Statistical Moments", MILCOM'86, pp.20.2.1–20.2.6.

[9] Ribble, C.S., "Automatic Modulation Recognition", Royal Military College, Kingston, Ontario, April 1987.

[10] Mammone, R.J., et al., "Estimation of Carrier Frequency, Modulation Type and Bit Rate of an Unknown Modulated Signal", Proc. of the ICC, 1987.

[11] Julius T. Tou and Rafael Gonzalez, "Pattern Recognition Principles", Addison–Wesley, 1974

[12] Richard P. Lippman, "Pattern Classification Using Neural Networks", IEEE Communications Magazine, November 1989, pp.47–64.

[13] Michèle Basseville, "Distance Measures for Signal Processing and Pattern Recognition", Signal Processing, December 1989, pp.349–369.

[14] Afifi, A.A. and Clark, Virginia, "Computer–Aided Multivariate Analysis", Lifetime Learning Publications, Belmont, California, 1984.

[15] K. Fukunaga, "Introduction to Statistical Pattern Recognition", Academic Press, 1972.

[16] Josef Kittler, "Mathematical Methods of Feature Selection in Pattern Recognition", International Journal of Man–Machine Studies, Vol.7, 1975, pp.609–637.

[17] R. A. Fisher, "The Use of Multiple Measurements in Taxonomic Problems", Annals of Eugenics, 7:179–188, 1936

[18] Tzay Y. Young, "Handbook of Pattern Recognition and Image Processing", Academic Press, 1986.

[19] Seymour Shlien, "Nonparametric Classification Using Matched Binary Decision Tree", to appear in Pattern Recognition Letters, 1990.

[20] Gerard T. LaVarnway, "An Introduction to CART: Classification and Regression Trees", in "Computing Science and Statistics", Proceeding of the 20th Symposium on the Interface, Fairfax, Virginia, April 1988, edited by E.J. Wegman et al..

[21] Watkins–Johnson Company, "WJ–8607 VHF/UHF Surveillance Receiver, Developmental Specification", Watkins–Johnson Company, Gaithersburg, Md., Change 2, January 1988.

[22] Winkler, C., "The Operation of an Advanced Modulation Recognition Algorithm", Technical Memorandum ERL–0185–TM, Electronics Research Laboratory, Defence Research Centre, Salisbury, Australia, February 1981.

[23] Calian Communications Systems Limited, "Feasibility Study for a HF/VHF ESM Receiver", Miller Communications Systems Limited, Kanata, Ontario, Canada, January 1978.

[24] Calian Communications Systems Limited, "Development of a Spectrum Surveillance Electronic Support Measures Receiver System–Operators Manual and Test Report", DSS Contract #2SR79–00069, MCS File #7829 PR/ce, Miller Communications Systems Limited, Kanata, Ontario, Canada, October 1980.

[25] Calian Communications Systems Limited, "The ESM Receiver Operator's Manual and Test Report", Miller Communications Systems Limited, Kanata, Ontario, Canada, January 1981.

[26] Calian Communications Systems Limited, "Frequency Resolution Algorithm Selection for Improved Identification Performance in the CAIR Receiver", MCS File No.8136, Miller Communications Systems Limited, Kanata, Ontario, Canada, November 1982.

[27] Calian Communications Systems Limited, "Operating Manual Intercept Receiver", CS File No.8136, Calian Communications Systems Limited, Kanata, Ontario, Canada, February 1983.

[28] Jondral, Friedrich, "Automatic Classification of High Frequency Signals", Signal Processing, Vol.9, 1985.

[29] Jondral, Friedrich, "Foundations of Automatic Modulation Classification", Stochastische Modelle und Methoden in der Informationstechnik, AEG Aktiengesellschaft, April 1989.

[30] Torrieri, D.J., "Principles of Military Communication Systems", Artech House, Massachusetts, 1981, p.11.

[31] Luiz, R., "HF/VHF Spectrum Surveillance ESM Receiver", Department of National Defence Canada, Ottawa, Technical Note No. 81–17, September 1981.

[32] Wakeman, Philip E.D., "Spectrum Surveillance Receiver System", US Patent 4.501.020, 19th February 1985.

[33] Gadbois, L.G. et al., "Identification of the Modulation Type of a Signal", IEEE International Conference on Acoustics, Speech, and Signal Processing, March 1985.

[34] Gadbois, L.G., "Modulation Identification", Royal Military College, Kingston, ONT., October 1985.

[35] Gadbois, L.G. et al., "Identification of the Modulation Type of a Signal", Signal Processing, Vol.16, 1989.

[36] UTL CANADA INC., "HFSS Modulation Recognition Study", DND/DEEM 4 Contract No. W8477–7–CC20/01–SV, August 1989.

[37] Weaver, C.S., et al., "The Automatic Classification of Modulation Types by Pattern Recognition", Technical Report No. 1829–2, Stanford Electronics Laboratories, Stanford University, Stanford, CA, April 1969.

[38] Callaghan, T.G., et al., "Sampling and Algorithms Aid Modulation Recognition", Microwave & RF, September 1985.

[39] Fabrizi, P.M., et al., "Receiver Recognition of Analogue Modulation Types", IERE Conference on Radio Receivers and Associated Systems, Bangor, Wales, 1986.

[40] Petrovic, P.M. et al., "An Automatic VHF Classifier" Mediterranean Electrotechnical Conference, MELECON'89.

[41] Aisbett, J., "Automatic Modulation Recognition Using Time Domain Parameters", Electronic Research Lab, Adelaide, Australia, Technical Report ERL–0367–TR, November 1986.

[42] Aisbett, Janet, "Automatic Modulation Recognition Using Time Domain Parameters", Signal Processing, Vol.13, No.3, pp.323–328, 1987.

[43] Vergara–Dominguez, L., Portillo–Garcia, J., et al., "Modulated Signal Classification: A General Approach", Proc. of the IASTED International Symposium, June 1987.

[44] Vergara–Dominguez, L., Portillo–Garcia, J., et al., "A Radio Communication Signal Classifier", EURASIP 1988, Signal Processing IV, pp.1361–1364, 1988.

[45] Adams, E.R. et al., "A Statistical Approach to Modulation Recognition for Communications Signal Monitoring and Surveillance", Fifth International Conference on Digital Processing of Signals in Communication, 1988.

[46] Tatsuoka, M., "Multivariate Analysis: Techniques for Educational and Psychological Research", Wiley, 1971.

[47] Richard P. Lippman, "An Introduction to Computing with Neural Nets", IEEE ASSP Magazine, April 1987.

[48] Ready, Patrick J. et al., "Modulation Detector and Classifier", US Patent 4,597,107, 24th June 1986.

[49] Kim, Kiseon and Polydoros, Andreas, "Modulation Detection/Classification for Phase-Modulated Digital Signals in Broadband Noise", Communication Sciences Institute, Departement of Electrical Engineering, University of Southern California, Los Angeles, November 1987.

[50] Kim, Kiseon and Polydoros, Andreas,"Digital Modulation Classification: The BPSK versus QPSK case", MILCOM, 1988.

[51] Gardner, William A. et al., "Cyclic Spectral Analysis for Signal Detection and Modulation Recognition", MILCOM, 1988.

[52] William A. Gardner, "Signal Interception: A Unifying Theorical Framework for Feature Detection", IEEE Transactions on Communications, Vol.36, No.8, August 1988, pp.897–906.

DOCUMENT CONTROL DATA

(Security classification of title, body of abstract and indexing annotation must be entered when the overall document is classified)

1. ORIGINATOR (the name and address of the organization preparing the document. Organizations for whom the document was prepared, e.g. Establishment sponsoring a contractor's report, or tasking agency, are entered in section 8.) NATIONAL DEFENCE DEFENCE RESEARCH ESTABLISHMENT OTTAWA SHIRLEY BAY, OTTAWA, ONTARIO K1A 0Z4 CANADA	2. SECURITY CLASSIFICATION (overall security classification of the document including special warning terms if applicable) UNCLASSIFIED

3. TITLE (the complete document title as indicated on the title page. Its classification should be indicated by the appropriate abbreviation (S,C or U) in parentheses after the title.)

MODULATION RECOGNITION: AN OVERVIEW (U)

4. AUTHORS (Last name, first name, middle initial)

LAMONTAGNE, RENE

5. DATE OF PUBLICATION (month and year of publication of document) MARCH 91	6a. NO. OF PAGES (total containing information. Include Annexes, Appendices, etc.) 68	6b NO OF REFS (total cited in document) 56

7 DESCRIPTIVE NOTES (the category of the document, e.g. technical report, technical note or memorandum. If appropriate, enter the type of report, e.g. interim, progress, summary, annual or final. Give the inclusive dates when a specific reporting period is covered.)

DREO TECHNICAL NOTE 91-3

8. SPONSORING ACTIVITY (the name of the department project office or laboratory sponsoring the research and development. include the address.)
NATIONAL DEFENCE
DEFENCE RESEARCH ESTABLISHMENT OTTAWA
SHIRLEY BAY, OTTAWA, ONTARIO K1A 0Z4 CANADA

9a. PROJECT OR GRANT NO. (if appropriate, the applicable research and development project or grant number under which the document was written. Please specify whether project or grant) 041LK11	9b. CONTRACT NO (if appropriate, the applicable number under which the document was written)
10a. ORIGINATOR'S DOCUMENT NUMBER (the official document number by which the document is identified by the originating activity. This number must be unique to this document.) DREO TECHNICAL NOTE 91-3	10b. OTHER DOCUMENT NOS. (Any other numbers which may be assigned this document either by the originator or by the sponsor)

11 DOCUMENT AVAILABILITY (any limitations on further dissemination of the document, other than those imposed by security classification)

(XX) Unlimited distribution
( ) Distribution limited to defence departments and defence contractors; further distribution only as approved
( ) Distribution limited to defence departments and Canadian defence contractors, further distribution only as approved
( ) Distribution limited to government departments and agencies; further distribution only as approved
( ) Distribution limited to defence departments; further distribution only as approved
( ) Other (please specify):

12 DOCUMENT ANNOUNCEMENT (any limitation to the bibliographic announcement of this document. This will normally correspond to the Document Availability (11). However, where further distribution (beyond the audience specified in 11) is possible, a wider announcement audience may be selected.)

13. **ABSTRACT** (a brief and factual summary of the document. It may also appear elsewhere in the body of the document itself. It is highly desirable that the abstract of classified documents be unclassified. Each paragraph of the abstract shall begin with an indication of the security classification of the information in the paragraph (unless the document itself is unclassified) represented as (S), (C), or (U). It is not necessary to include here abstracts in both offical languages unless the text is bilingual).

THE INTEREST FOR A SYSTEM ABLE TO AUTOMATICALLY IDENTIFY THE MODULATION TYPE OF AN INTERCEPTED RADIO SIGNAL IS INCREASINGLY EVIDENT FOR MILITARY AND CIVILIAN PURPOSES. ALTHOUGH IN THE PAST SOME AUTHORS LOOKED AT THE PROBLEM, NOBODY FOUND "THE SOLUTION" AND THE PROBLEM REMAINS. THEREFORE, AN OVERVIEW OF THE PROPOSED TECHNIQUES IS USEFUL IN ORDER TO ASSESS THE ACTUAL SITUATION.

THIS DOCUMENT PRESENTS CLASSIFICATION TECHNIQUES AND FEATURES (PARAMETERS CHARACTERIZING THE MODULATION TYPES) USED FOR MODULATION RECOGNITION.

QUE CE SOIT POUR DES FINS MILITAIRES OU CIVILES, L'INTERÊT D'UNE MACHINE CAPABLE D'IDENTIFIER AUTOMATIQUEMENT LE TYPE DE MODULATION D'UN SIGNAL INCONNU EST EVIDENT. BIEN QUE PAR LE PASSE CERTAINS AUTEURS SE SOIENT PENCHES SUR LE PROBLEME, PERSONNE N'A OBTENUE D'ECLATANT RESULTATS INDIQUANT LA ROUTE A SUIVRE. PAR CONSEQUENT IL N'Y A PAS LIEU DE CONCENTRER SES EFFORTS SUR UN SEUL AUTEUR MAIS SUR L'ENSEMBLE DES TECHNIQUES PROPOSEES.

CE DOCUMENT COUVRE DONC LE SUJET D'UNE FACON GLOBALE, PRESENTANT LES TECHNIQUES POUR CLASSIFICATION AINSI QUE LES PARAMETRES UTILISES POUR CARACTERISER ET DISTINGUER LES DIVERS TYPES DE MODULATION, TEL QUE PROPOSES PAR LES AUTEURS.

14. KEYWORDS, DESCRIPTORS or IDENTIFIERS (technically meaningful terms or short phrases that characterize a document and could be helpful in cataloguing the document. They should be selected so that no security classification is required. Identifiers, such as equipment model designation, trade name, military project code name, geographic location may also be included. If possible keywords should be selected from a published thesaurus, e.g. Thesaurus of Engineering and Scientific Terms (TEST) and that thesaurus-identified. If it is not possible to select indexing terms which are Unclassified, the classification of each should be indicated as with the title.)

MODULATION RECOGNITION
MODULATION CLASSIFICATION
MODULATION IDENTIFICATION